Relations Among Performance and Symptom Validity, Mild Traumatic Brain Injury, and Posttraumatic Stress Disorder Symptom Burden in Postdeployment Veterans

Robert D. Shura[1,2,3] · Ruth E. Yoash-Gantz[1,2] · Treven C. Pickett[4] · Scott D. McDonald[5,6,7] · Larry A. Tupler[1,8,9]

Abstract

The purpose of this study was to evaluate relationships among the Word Memory Test (WMT), symptom validity test (SVT) indices of the Personality Assessment Inventory (PAI), history of mild traumatic brain injury (mTBI), and posttraumatic stress disorder (PTSD) symptom burden. Participants were postdeployment, predominantly male (88.5%) veterans ($N=417$) who completed a neurocognition study that included the WMT and PAI. Correlations, chi-square analyses, and ANOVAs were used for analyses. Results of aim 1 examining relations among the two tests found that 20.4% produced invalid scores on the WMT (regardless of PAI scores), 13.8% produced an invalid PAI (regardless of WMT scores), and 4.6% were invalid on both tests. Of the 4 original PAI validity scales, only Negative Impression Management was related to WMT failure; the supplementary Malingering Index was also significant at a smaller effect size. The second aim evaluated mTBI and PTSD symptoms in relation to validity scores. History of mTBI was associated with invalid WMT scores but not PAI scores; follow-up analyses indicated that injuries sustained during deployment were significantly more likely to produce invalid WMT scores than non-deployment injuries. Contrary to hypotheses, PTSD symptom burden was related to WMT failure but not overreporting on the PAI. After dividing the sample into four groups based on valid versus invalid WMT and PAI scores, the invalid PVT valid SVT group had the highest proportion of mTBI, whereas PTSD symptom burden was highest in the groups that invalidated both measures or only the WMT. Service-connected disability status was unrelated to either type of validity. Given invalidating both types of validity measures is uncommon, the importance of sampling both types of validity is highlighted.

The importance of assessing performance validity (fidelity of performance to genuine ability on neurocognitive tests) and symptom validity (accuracy of self-report) in neuropsychological examinations has become increasingly recognized in the past two decades (Bush et al., 2005;

Heilbronner et al., 2009), and up to 25% of papers in two popular neuropsychology journals have been devoted to the topic (Martin et al., 2015). In military and veteran samples, nearly one in three will fail standalone performance validity tests (PVTs; Denning & Shura, 2019), highlighting the

✉ Robert D. Shura
Robert.Shura2@va.gov

1 Mid-Atlantic Mental Illness Research, Education, and Clinical Center, NC, Durham, USA

2 Research & Academic Affairs Service Line, Salisbury Veterans Affairs Medical Center, Salisbury, NC, USA

3 Department of Neurology, Wake Forest School of Medicine, Winston-Salem, NC, USA

4 National Intrepid Center of Excellence, Walter Reed National Military Medical Center, Bathesda, MD, USA

5 Central Virginia Veterans Affairs Health Care System, Richmond, VA, USA

6 Department of Physical Medicine and Rehabilitation, Virginia Commonwealth University, Richmond, VA, USA

7 Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA

8 Durham Veterans Affairs Medical Center, Durham, NC, USA

9 Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA

importance of including PVTs in neuropsychological evaluations. Likewise, interpretation of multi-scale, self-report measures begins with interpretation of symptom validity indices, and a recent study of over 17,000 Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) protocols in the Department of Veterans Affairs (VA) found invalidity rates from 5% to 27% depending on scale (Ingram et al., 2019). A number of studies indicate that a subgroup of those who invalidate PVTs may also invalidate self-report symptom validity tests (SVTs). Although PVT and SVT presentations may be mostly independent, research has suggested that they are not mutually exclusive. The purpose of this paper is to evaluate patterns of invalidity across the Word Memory Test (WMT) and Personality Assessment Inventory (PAI) validity scales as well as the relations among PVT and SVT measures to a history of mild traumatic brain injury (mTBI) and current posttraumatic stress disorder (PTSD) symptom burden.

Numerous standalone PVTs are available to assess validity within several cognitive domains. The WMT (Green, 2005) is a forced-choice, memory-based PVT that is commonly used by neuropsychologists and rated as one of the most accurate PVTs available (Sharland & Gfeller, 2007). Some studies have shown that the test is easy enough to be passed by those with significant cognitive impairment (Carone, 2014; Green & Flaro, 2015; Green et al., 2002), and increased interpretive ability is achievable using profile analysis, such as for those with dementia (Green et al., 2011). In one meta-analysis (Sollman & Berry, 2011), the WMT produced the second highest sensitivity (75.1%), after only the Victoria Symptom Validity Test (81.5%), though with a reduced specificity. In veteran and military samples, the WMT was the second most frequently studied standalone PVT (19 studies), after only the Test of Memory Malingering (TOMM; Denning & Shura, 2019). Of note, the WMT had a mean failure rate of 39% across veteran and military studies, and although some have suggested the measure is associated with a high number of false positives, this assertion has been debated in the literature (Bowden et al., 2006; Erdodi et al., 2019; Green et al., 2009; Greve et al., 2008). For example, Erdodi and colleagues (2019) dispute the false-positive allegation on logical and empirical grounds: failure rate on the WMT in their study was slightly higher than for other measures, but those who failed showed independent evidence of invalid performance. Regardless, given the generally high sensitivity and established common use in veteran and military studies, the WMT is a suitable PVT to evaluate validity issues in veteran samples.

The PAI includes four original, primary symptom-validity scales: two scales that measure inconsistency (Inconsistency [ICN], Infrequency [INF]), an underreporting scale (Positive Impression Management [PIM]), and an overreporting scale (Negative Impression Management [NIM]). Additionally,

several supplementary validity indices are available, including the Malingering Index (MAL), Rogers Discriminant Function (RDF), Defensiveness Index, and Cashel Discriminant Function (Morey, 2007). PAI validity scales and indices have shown promise in specifically identifying overreporting across a number of populations, such as forensic psychiatric inpatients (Wang et al., 1997), criminal defendants (Kucharski et al., 2007), compensation-seeking mTBI patients (Whiteside et al., 2012), and Medical Evaluation Board exams (Mooney et al., 2018). Prior studies have also shown these PAI scales to be useful in detecting exaggeration of specific psychiatric disorders, such as PTSD (Calhoun et al., 2000; Thomas et al., 2012), major depressive disorder, generalized anxiety disorder, and schizophrenia (Hopwood et al., 2007). Like the WMT, the PAI is in common usage among VA psychologists.

Although sophisticated feigners may be informed enough to adjust presentation in a way that is less obvious or focused on specific domains (Nelson et al., 2003), less-sophisticated approaches may involve global exaggeration. Similarly, outside of intentional feigning, those with "poor effort" may also either put forth variable effort across an exam, or globally not engage. In a forensic sample, an exploratory factor analysis of three PVTs and a number of Minnesota Multiphasic Personality Inventory-2 (MMPI-2) scales identified distinct cognitive (*cognitive effort*) and psychological validity factors (*underreporting*, *overreporting neurotic symptoms*, and *overreporting psychotic symptoms*; Nelson et al., 2007). This finding suggests that different invalid presentations exist. Additional studies support the concept of performance validity and symptom validity as distinct factors (Copeland et al., 2016; Jurick et al., 2019; McCaffrey et al., 2003; Van Dyke et al., 2013). However, a subgroup has been identified in the literature that invalidates both PVTs and SVTs (Armistead-Jehle & Buican, 2012; Larrabee, 2003a, 2003b; Whiteside et al., 2009). Due to consistent and global exaggeration across multiple domains, this subgroup might be more likely to include intentional exaggerators and unsophisticated feigners, as opposed to those who invalidate testing for other reasons.

A number of studies have examined PAI validity scales in conjunction with failed PVT performance. Studies suggest that the validity scales of the PAI are useful compared with invalidity on the TOMM (Aikman & Souheaver, 2008; Whiteside et al., 2009), the Victoria Symptom Validity Test (Haggerty et al., 2007), and the Dot Counting Test (Sumanti et al., 2006). Two studies were found examining the PAI validity scales in relation to the WMT, both using a sample of active military service members with a history of TBI (Armistead-Jehle & Buican, 2012; Lange et al., 2012). In comparing those with valid versus invalid scores on the WMT in a mixed clinical/forensic sample, Armistead-Jehle and Buican (2012) found that those who failed had significantly higher scores on the PAI NIM ($d = 0.49$) and

MAL ($d = 0.35$), with significantly lower scores on the PIM ($d = 0.40$). The other PAI validity scores (INC, INF, RDF) were not significantly different between groups. Of note, context mattered; differences became non-significant in the forensic subsample, though the clinical subsample continued to see significant elevations on NIM ($d = 0.60$) and MAL ($d = 0.45$) only. Lange and colleagues (2012), using the four primary SVTs of the PAI, reported that service members with mTBI failing the WMT had significantly higher NIM scores than both mild and severe TBI patients passing the WMT, though INC, INF, and PIM were not significantly different. In sum, limited research has indicated a relationship between invalid scores on the WMT and symptom overreporting, specifically for NIM and MAL, on the PAI. These relations among validity indicators are important to consider in view of some of the more common psychiatric presentations (e.g., PTSD) of veterans of operations Enduring Freedom, Iraqi Freedom, and New Dawn (OEF/OIF/OND).

Beyond psychometric relationships across PVTs and SVTs, certain conditions may additionally relate to performance and symptom invalidity. PTSD is one of the most common psychiatric disorders for the postdeployed veteran cohort (Kessler et al., 2014)t, with one meta-analysis finding a prevalence of 23% in returning veterans (Fulton et al., 2015). The mTBI population is also particularly relevant to the recent postdeployment cohort (Kessler et al., 2014), with over 400,000 service members diagnosed with TBI injuries from 2000 to 2019, 83% of which were mild in severity (Defense and Veterans Brain Injury Center, 2019). Additionally, PTSD is commonly comorbid with a history of TBI (Dismuke-Greer et al., 2019; Hoge et al., 2008; Lew et al., 2009; Pugh et al., 2014). Moreover, in fiscal year 2018, PTSD was the most common psychiatric disability condition at 59%, far higher than the second most-common condition (depression) at 12% (Veterans Benefits Administration, 2019). Similarly, in individuals who have sustained an mTBI, complete recovery of postconcussive cognitive and psychiatric symptoms is expected by 3 months, with litigation and disability factors cited as significantly related to persisting complaints (Belanger et al., 2005; Berry et al., 1995; Carroll et al., 2004; Panayiotou et al., 2010; Youngjohn et al., 1997). In veteran samples, the influence of service-connected disability may reach beyond the direct forensic exam, and invalidity related to disability concerns potentially present even in non-forensic clinical exams. Thus, it is critical to consider the possible influence of overreporting and underperformance on symptom portrayal when evaluating both mTBI and PTSD. Important implications attend both clinical (Veterans Health Administration) and programmatic (Veterans Benefits Administration) demands due to the frequency of the association of these conditions with service-connected disability, as invalid presentations could lead to extensive, inappropriate financial loss.

In addition to service-connected disability issues that can occur following mTBI, other factors may also affect outcomes in veteran samples with a history of mTBI. Relevant to military and veteran populations are differences between injuries sustained in a combat environment versus those in the civilian context. Recent studies of post-deployed veterans have found poorer outcomes following deployment mTBI compared with non-deployment mTBI, including for psychiatric diagnoses and symptoms, sleep, alcohol use, pain severity, and cognitive functioning (Martindale et al., 2018, 2020). Thus, injury context represents an important variable to address when analyzing mTBI relationships in analyses.

Given that Armistead-Jehle and Buican (2012) found significant differences in NIM, PIM, and MAL across those passing versus failing the WMT in their sample of 333 service members, the first aim of the current study is to replicate the PAI/WMT relationship in a similarly sized sample in a different evaluation context (primarily research versus clinical/forensic). For the first hypothesis, given base rates in prior studies, we predicted that (H1) PVT and SVT rates will occur in the following order: pass both > fail either PVT or SVT > fail both. As a follow-on second hypothesis, we predicted that (H2) the presence of NIM failure will enhance the odds of also observing PVT failure more so than the presence of PVT failure will enhance the odds of observing SVT failure. Third, we hypothesized that (H3) the relationship between PVT failure and SVT failure will only apply to the NIM and MAL scales. That is, associations between PVT failure and SVT failure will be significant for overreporting, but not underreporting or non-content responding, similar to prior research (Armistead-Jehle & Buican, 2012; Lange et al., 2012).

Using the four primary PAI SVT scales in a sample of service members, Lange and colleagues (2012) found that only NIM was significantly higher in an mTBI group that failed the WMT, compared with mild and severe TBI groups that passed the WMT. Therefore, the second aim of this study was to examine similar relations among the WMT, PAI SVT scales, and veterans with a history of mTBI, as well as to incorporate PTSD symptom burden given the high rate of disability for PTSD and relevance of both mTBI and PTSD to the postdeployment population. Given that mTBI and PTSD symptoms reflect neurological versus psychiatric constructs, respectively, we expect a double dissociation in relation to PVT and SVT status. Therefore, we hypothesized that (H4) PVT but not SVT failure will be significantly associated with a history of mTBI. As follow-up analyses, and given prior research, we predicted higher effect sizes for deployment mTBI compared with non-deployment mTBI. Similarly, we hypothesized that (H5) SVT (NIM and MAL) but not PVT (WMT) invalidity will be significantly associated with PTSD symptom burden (PAI Traumatic Stress [ARD-T]),

further establishing independence of the validity constructs. Finally, in lieu of the potentially salient legal and other translational implications, we perform a post hoc, exploratory examination of the relevance of service-connection status to both PVT and SVT outcomes: we evaluate both status of already receiving service connection and plan to apply for disability.

Method

Participants

The present analyses were conducted on a sample of predominantly postdeployed military personnel serving in Iraq or Afghanistan. Candidate participants consisted of 432 OEF/OIF/OND-era active-duty and veteran service members assessed in a multi-site protocol conducted by the Neurocognition Laboratory of the VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center. Institutional Review Board approval was obtained at each of three VA Medical Center (VAMC) sites located in the Mid-Atlantic region. All participants voluntarily supplied verbal and written informed consent prior to any study activities. Participants were identified and recruited from the Post-Deployment Mental Health study (PDMH; Brancu et al., 2017), a VAMC Polytrauma/TBI System of Care inpatient, residential, and outpatient rehabilitation program; and the wider multi-site VAMC patient population. Additional information on a subset of these data can be found in McCormick and colleagues (2013).

Subjects were heterogeneous for demographics and diagnoses present, though all served on or after September 11, 2001. Formal *DSM-IV-TR* diagnostic categorization was not assessed on the day of cognitive testing; however, *DSM-IV-TR* criteria were evaluated in those who completed the entry PDMH study but only used for eligibility criteria given the varying times participants completed the two studies. Specific study exclusion criteria included completion of neuropsychological evaluation during the prior 6 months, active psychotic symptoms, and presence of an active substance-use disorder. Individuals were reimbursed for their time and travel. The welfare of human subjects was protected. Demographic variables for the sample are presented in Table 1. The sample was generally male (88.5%) and White (65.7%), with a mean age of 35.0 years and at least some college attendance. It is unlikely that individuals with major neurodegenerative disorders were included in this study due to the sample characteristics and eligibility criteria. Additionally, 14.6% of the sample was still on active duty, thus ineligible for disability, and 50.5% of the remaining sample was receiving service-connected disability for any reason (range 0–100%). As part of the consent process, participants were told that results would not be available to disability examiners nor used for disability purposes; however, a small

Table 1 Descriptive statistics for sample (*N*=417)

Variable	*M* or *n*	*SD* or %	Min–max
Age	35.0	9.8	19–64
Education			
Incomplete high school	1	<1%	
GED	7	1.7%	
High school	124	29.7%	
Some college	126	30.2%	
Associate degree	56	13.4%	
College degree	80	19.2%	
Graduate degree	23	5.5%	
Male	369	88.5%	
Ethnicity			
Non-Hispanic			
White	262	62.8%	
African American	123	29.5%	
Asian	4	1.0%	
Other	11	2.6%	
Hispanic			
White	12	2.9%	
African American	1	0.2%	
Other	4	1.0%	
Service connection (*n*=410)			
Yes	207	50.5%	
No	143	34.9%	
Active duty	60	14.6%	
Service connection percentage (*n*=201)	49.1	29.7	0–170
Plan to apply for disability (*n*=412)	35	8.5%	
Traumatic brain injury			
Mild	72	17.2%	
Moderate/severe	67	16.1%	
Missing details	13	3.1%	
Personality Assessment Inventory (*n*=412)			
≥ 1 scale failure	57	13.8%	
Inconsistency (*n*=411)			
T score	52.1	10.0	34–82
Failed	17	4.1%	
Infrequency			
T score	52.2	9.7	40–90
Failed	14	3.4%	
Positive Impression Management			
T score	47.2	12.3	15–77
Failed	17	4.1%	
Negative Impression Management			
T score	56.2	14.6	44–125
Failed	15	3.6%	
Malingering Index			
Score	0.76	0.97	0–6
Failed	17	4.3%	
Traumatic stress T-score	63.1	18.5	41–99

Table 1 (continued)

Variable	M or n	SD or %	Min–max
Word Memory Test ($n=398$)			
Invalid on ≥ 1 "effort" scale	85	21.4%	
Immediate Recognition			
T score	92.8	11.0	22.5–100
Failed	55	13.8%	
Delayed Recognition			
T score	92.7	11.9	25.0–100
Failed	52	13.1%	
Consistency			
T score	90.5	11.7	42.5–100
Failed	82	20.6%	

number provided dual consent to use the data for clinical purposes as well.

Measures

A standardized, fixed battery of neurocognitive, psychological, performance validity, and self-report symptom measures was administered. The larger NeuroCognition study from which these data derive involves a variety of aims related to neurocognitive functioning of postdeployment veterans. Briefly, the study included 21 cognitive measures, the WMT, and 5 self-report measures (see McCormick et al., 2013 for additional description). Tests were administered by neuropsychologists, postdoctoral fellows under supervision of neuropsychologists, and trained and supervised psychology technicians. The WMT and PAI were administered on identical Dell Latitude D410 laptop computers at each of the three VAMC sites.

The WMT was selected as the primary PVT for the larger study given that it is in common use by VA neuropsychologists, thus familiar to the field; it shows high sensitivity, and the computerized presentation minimizes administration error. The WMT (Green, 2005) presents a series of 20 paired words for two learning trials followed by a forced-choice Immediate Recognition (IR) trial and subsequent Delayed Recognition (DR) trial after 30 min. A Consistency Score (CNS) is also calculated from the number of items that were correct on both IR and DR trials compared with correct on one trial and incorrect on the other trial. These three scores are the primary "effort" scores for the measure, and as such were used for analyses. The WMT was scored using the established cutoffs to identify valid and invalid scores (Green, 2005). For this study, if at least one of the three easy ("effort") scores was in the invalid score range, the participant was identified as a WMT fail. Participants whose scores were all in the valid score range were identified as a WMT pass. When designing the battery, not all WMT subtests were included given the objectives, time

demands, and logistic constraints of the study. The Genuine Memory Impairment Profile ("dementia profile") cannot be examined.

The PAI is a 344-item, multi-scale, self-report measure of psychiatric distress (Morey, 2007). The instrument contains four main SVTs. The INC scale consists of 10 pairs of similar items, half that should be endorsed differently and half that should be endorsed similarly, thus measuring inconsistent responding. The INF scale consists of 12 items that should be infrequently endorsed by the general population, and measures careless or random responding. PIM contains 12 items and measures symptom underreporting. NIM includes 12 items across two question types: items rarely endorsed in clinical samples and items developed to detect feigned and exaggerated symptoms. PTSD symptom burden was also measured using the PAI ARD-T scale, a subscale of the Anxiety-Related Disorders clinical scale; the 8 items measure symptoms common following traumatic events, such as re-experiencing (Morey, 2007). Of note, ARD-T items are non-overlapping with the NIM items. A secondary scale was also evaluated: MAL is an index from 0 to 8 with points based on scale cutoffs and comparisons; this scale was included as, in addition to the NIM, it was significant in Armistead-Jehle and Buican's (2012) active-duty clinical sample.

The presence of a history of TBI was established using a self-report measure presented on computer modified from a questionnaire previously published (Ivins et al., 2003), either typed in by the participant during the PDMH study ($n=357$, 85.6%) or completed by research staff based on the medical records from the Polytrauma rehabilitation site ($n=60$, 14.4%). The respondent completing the questionnaire enters information for up to 6 injuries including date of injury, duration of loss of consciousness, posttraumatic amnesia, and alteration of consciousness. Using this information and VA/Department of Defense TBI criteria, participants were determined to have no injuries, one or more mild injuries, or 1 or more moderate/severe injuries. Those with a history of moderate or severe TBI ($n=63$) were excluded from the TBI analyses, as the TBI hypothesis focused on mTBI only. Injury events were also classified as to whether or not the injury occurred during deployment, based on the injury dates provided on the TBI measure and self-reported dates of all combat deployments. The data for deployments, service-connection status, and plan to apply for disability were obtained from either the prior PDMH study visit or medical records.

Results

All statistical analyses were computed using SAS version 9.4 (SAS Institute Inc., Cary, NC). Demographic variables were examined using univariate descriptive statistics. A

significance level of 0.05 was set *a priori* for all inferential tests. The sample was drawn from the 432 candidate participants, from which 15 participants missing both the WMT and the PAI due to technical difficulties were excluded, leaving a final $N = 417$ for analyses: 24 were missing either the WMT or PAI, leaving $n = 393$ for analyses requiring both. In those with WMT data ($n = 398$), 85 (21.4%) failed the WMT, 55 failed IR, 52 failed DR, and 82 failed CNS (see Table 1). In those with PAI data ($n = 412$), 57 (13.8%) failed at least one of the primary SVT scales, 6 of whom failed 2 scales: 15 failed NIM, 17 failed PIM, 17 failed ICN, and 14 failed INF. Of the 6 individuals that failed 2 SVT scales, 4 failed both ICN and INF, 1 failed both PIM and INF, and 1 failed both NIM and ICN. Additionally, 17 (4.1%) produced invalid scores on the supplementary scale MAL. For mTBI analyses, the sample was $n = 337$ after excluding those with missing data or a history of moderate-to-severe TBI.

The first hypothesis predicted that base rates would be present in the following order: pass both > fail either PVT or SVT > fail both. Using dichotomous pass-fail variables for producing an invalid score on any of the three WMT PVT subtests or any of the four primary PAI SVT scales, the hypothesis was supported. The chi square was significant, χ^2 (1, $N = 393$) = 5.50, $p = 0.019$, $\phi = 0.12$, with the following percentages: pass both $n = 273$ (69.5%), fail either PVT or SVT $n = 102$ (26.0%), and fail both $n = 18$ (4.6%). Second, we hypothesized that the presence of invalid SVT scores would enhance the odds of also observing invalid PVT scores more so than the presence of invalid PVT scores would enhance the odds of observing invalid SVT scores. Of those failing the PAI, 34.0% also failed the WMT. Of those failing the WMT, 21.2% also failed the PAI. However, a Z-test of proportions was not significant, z = 1.66, $p = 0.098$, $\phi = 0.01$. Third, we hypothesized that an invalid score on NIM specifically would relate to

invalid WMT performance. This prediction was also supported, as of the four primary SVT scales, only elevations on NIM were associated with a significantly different percentage of failure on the WMT (see Table 2). Follow-up analysis evaluating the MAL was also significant, but at a smaller effect size than NIM. Therefore, non-content-related response bias (e.g., random responding) and underreporting were not related to underperformance.

The second aim of this study was to evaluate the relationship between invalidity and two common postdeployment conditions: history of mTBI, a primarily neurologic diagnosis, and PTSD symptom burden, related to a psychiatric diagnosis. The fourth hypothesis predicted that a history of mTBI would relate to PVT more so than SVT performance. For these analyses, we excluded all moderate/severe TBI ($n = 63$) and those missing information concerning one or more TBIs ($n = 14$), leaving a final mTBI-related sample of $n = 327$: 233 (69.1%) had no history of mTBI, 72 (21.4%) had a history of 1 mTBI, and 32 (9.5%) had 2 or more mTBIs. This hypothesis was supported: a higher percentage of invalid WMT scores was seen in those with a history of mTBI (29.3%) compared with those without a history of mTBI (12.95%), χ^2 (1, $N = 323$) = 12.45, $p < 0.001$, $\phi = 0.20$. The same findings were not apparent for invalidity rates on NIM: those with a history of mTBI invalidated at a rate of 2.94% compared with the nomTBI rate of 3.03%, χ^2 (1, $N = 333$) = 0.00, $p = 0.965$, $\phi = -0.00$.

The sample was divided into four groups: valid scores on both PVT and SVT measures ($n = 233$), valid PVT but invalid SVT ($n = 28$), invalid PVT but valid SVT performance ($n = 49$), and invalid on both measures ($n = 9$; see Table 3 for brief descriptives). A chi-square analysis comparing proportions with a history of mTBI versus no mTBI history across the four groups was significant, χ^2 (1, $N = 319$) = 14.81, $p = 0.002$, $\phi = 0.22$. Proportions comparing mTBI with no TBI were relatively the same in

Table 2 Comparison of performance and symptom validity failure rates ($n = 393$)

Symptom validity scale	Status	WMT Pass, $n = 308$	WMT Fail, $n = 85$	χ^2	Fisher's exact test[a] or p	ϕ
NIM	Pass	301 (76.6%)	77 (19.6%)	**9.25**	**.002**	**0.15**
	Fail	7 (1.8%)	8 (2.0%)			
PIM	Pass	297 (75.6%)	80 (20.4%)	0.91	.340	0.05
	Fail	11 (2.8%)	5 (1.3%)			
INF	Pass	300 (76.3%)	82 (20.9%)		.710[a]	0.02
	Fail	8 (2.0%)	3 (0.8%)			
ICN	Pass	296 (75.5%)	81 (20.7%)		1.000[a]	−0.01
	Fail	12 (3.1%)	3 (0.8%)			
MAL	Pass	298 (75.8%)	78 (19.9%)	**4.01**	**.045**	**0.10**
	Fail	10 (2.5%)	7 (1.8%)			

Bold scores are statistically significant

WMT Word Memory Test, *NIM* Negative Impression Management, *PIM* Positive Impression Management, *INF* Infrequency, *ICN* Inconsistency, *MAL* Malingering Index

the invalid PVT and invalid SVT group (3.1% versus 2.7%) and the valid PVT invalid SVT group (9.3% versus 8.6%). In contrast, the valid PVT valid SVT group had a significantly lower proportion of participants with (60.8%) than without (78.4%) mTBI. This effect appeared to be driven by the invalid PVT valid SVT group, with a higher proportion of mTBI (26.8%) than no mTBI (10.4%).

Given that differential outcomes have been observed across the difference in context of combat deployment compared with non-deployment (Martindale et al., 2018, 2020), we tested whether having an mTBI in the deployment context ($n = 65$) versus solely in a non-deployment context ($n = 43$) led to differential rates of invalid PVT scores. This finding was the case: 25 (38.5%) of those with deployment mTBI failed the PVT versus 6 (14.0%) of those with solely non-deployment mTBI failed the PVT, χ^2 (1, $N = 108$) $= 7.60$, $p = 0.006$, $\phi = 0.265$.

Finally, we hypothesized (H5) the converse result, i.e., that PTSD symptom burden would be related to SVT failure but not PVT failure. Correlations indicated that invalid status on both validity indices was significantly related to PTSD (ARD-T): WMT, $r = -0.26$, $p < 0.001$; NIM, $r = -0.28$, $p = 0.004$, and that PTSD symptom burden showed a similar, significant relationship to both invalid PVT and invalid SVT outcomes ($z = 0.30$, $p = 0.761$). Next, an ANOVA examined differences in mean ARD-T scores between the four validity groups. Results were significant, F (3, 392) $= 11.37$, $p < 0.001$, $\eta^2 = 0.09$. Tukey's pairwise comparisons revealed two significant group differences: valid PVT valid SVT compared with invalid PVT invalid SVT (mean difference $= 18.1$) and PVT valid SVT invalid (mean difference $= 11.0$). Table 3 presents group means. This hypothesis was not supported, as PTSD symptom burden was not related to SVT status only.

As a post hoc analysis, service-connected disability was also evaluated. Those who failed the WMT did not show a significantly higher proportion of having service connection compared with those who passed, χ^2 (1, $N = 391$) $= 0.22$, $p = 0.638$, $\phi = 0.024$. Similarly, those who failed NIM or MAL also did not show a significantly higher proportion of having service connection compared with those who passed:

Table 3 Descriptive comparisons across performance and symptom validity status groupings ($n = 319$)

Validity group	No mTBI, n (%)	mTBI, n (%)	ARD-T, M (SD)
PVT (−) SVT (−)	6 (66.7%)	3 (33.3%)	78.1 (22.5)
PVT (−) SVT (+)	23 (46.9%)	26 (53.1%)	71.0 (17.9)
PVT (+) SVT (−)	19 (67.9%)	9 (32.1%)	66.1 (20.7)
PVT (+) SVT (+)	174 (74.7%)	59 (25.3%)	60.0 (17.2)

(−) signifies an invalid score and (+) a valid score

mTBI mild traumatic brain injury, *ARD-T* Personality Assessment Inventory Traumatic Stress scale, *PVT* performance validity test, *SVT* symptom validity test

NIM Fisher's exact test $p = 0.597$, $\phi = 0.029$; MAL, χ^2 (1, $N = 405$) $= 1.94$, $p = 0.164$, $\phi = -0.069$. Additionally, no differences were observed in the proportion of participants with invalid PVT or SVT scores for those positive versus negative for the intent to apply for disability: invalid WMT, χ^2 (1, $N = 393$) $= 0.46$, $p = 0.500$, $\phi = 0.034$; invalid NIM Fisher's exact test $p = 0.373$, $\phi = 0.013$; invalid MAL Fisher's exact test $p = 0.342$, $\phi = 0.023$. In other words, neither presence of established disability nor intent to apply for disability exhibited a relationship to failure on either the WMT or PAI NIM/MAL.

Discussion

The first aim of this study was to evaluate the relationship between performance and symptom validity test failure, as prior research has suggested that the two constructs are independent but not mutually exclusive. Consistent with our first hypothesis, this study found that although performance and symptom validity are separate constructs, a small group of participants invalidate both types of tests. Failing both validity measures, however, was uncommon: 85 participants (20.4%) failed the PVT, 53 participants (13.8%) failed the SVT, and 18 participants (4.6%) failed both. These failure rates are generally consistent with rates reported in prior studies. For example, 11.2% of a non-veteran sample failed both the TOMM and RBS scale of the MMPI-2/RF in one study (Copeland et al., 2016), and 4% failed both PVTs and SVTs in another study with a veteran sample using a variety of measures (TOMM, Reliable Digit Span, Medical Symptom Validity Test, and MMPI-2 scales; Van Dyke et al., 2013).

The relationship between PVT failure and SVT failure was generally driven by overreporting, as only NIM and MAL were related to underperformance. This finding suggests that those responding carelessly or in an indiscriminate manner (i.e., process-based styles captured by non-content validity scales) would not necessarily invalidate cognitive testing. In other words, an individual responding randomly to a questionnaire is unlikely to also approach performance-based tests in a random manner. Additionally, those responding to the PAI in a defensive manner did not tend to also invalidate the WMT. This finding is somewhat intuitive, as an individual presenting in an overly positive manner is also likely to be motivated to perform well on cognitive tests. Given that the subgroup that failed both PVT and SVT measures was overreporting on the SVT, this particular group may reflect a subtype of examinees demonstrating invalid performance that is "pan-symptomatic": overreporting distress and acting behaviorally on testing in a manner that invalidates both protocols. If intentional, this presentation may reflect an unsophisticated approach to malingering,

though of note, this is an uncommon presentation in this sample. If unintentional, this presentation may reflect lack of insight or possibly a form of generalized uncooperativeness with the testing situation.

Next is the issue of directionality, which we hypothesized would show that SVT failure would exhibit a greater association with PVT failure than the reverse. Although not statistically significant, results were in the hypothesized direction, as about 1 in 3 who invalidated the PAI also failed the WMT, compared with approximately 1 in 5 who failed the WMT also failed the PAI. This result is consistent with the presence in many psychiatric disorders of impaired cognition as a symptom (e.g., decreased concentration in PTSD), though emotional symptoms may be less typically associated with cognitive disorders by the lay public. In forensic contexts, those feigning psychiatric disorders thus might be likely to fail PVTs due to beliefs that decreased cognition is part of the syndrome. However, those feigning cognitive issues (e.g., TBI litigants) may focus on cognitive symptoms only, or even downplay emotional issues. Additional research may inform this hypothesis, as our findings were only suggestive and not formally statistically significant in this nonforensic sample. Lastly, we hypothesized that the relationship between PVT performance and the PAI SVTs would be driven by NIM, based on prior literature (Armistead-Jehle & Buican, 2012; Mooney et al., 2018). Results support the idea that both NIM/MAL and failing a performance test are forms of presenting poorly, or biased response in the deviant direction. Thus, the relationship found between SVT and PVT failure is one between underperformance and overreporting specifically.

The second aim of this study was to explore a double dissociation reflected in findings that invalid performance validity would relate to a history of mTBI (neurological injury), whereas current PTSD symptoms (psychiatric status) would be related to overreporting but not invalid performance. Of note, although a history of mTBI was related solely to performance validity, PVT status was also more related to PTSD symptom burden than SVT status. Those with a history of sustaining an mTBI were more likely to produce invalid scores on the WMT but not on NIM compared with those without a history of mTBI. Further, from a follow-up analysis, of those with mTBI, those with an mTBI sustained in a deployment context were significantly more likely to invalidate the WMT than those with exclusively a non-deployment mTBI history. In a prior study using a subsample of the current full sample, context of testing (research versus dual research/clinical) was found to affect WMT failure rates (McCormick et al., 2013). It is possible that the effects of context may extend from implications of an evaluation to the situation in which an injury occurs (deployment versus non-deployment) and similarly affect validity, even when the context was in the remote past. These context-based

behaviors appear even outside of situations contextualized by secondary-gain issues, given that our sample was non-forensic.

Although PTSD symptoms were equally correlated with PVT and SVT statuses, contrary to our fifth hypothesis, PTSD symptom burden was primarily related to PVT status, with a significant effect seen for highest scores in the invalid PVT invalid SVT group and the invalid PVT valid SVT group, both compared with the valid PVT valid SVT group. Our finding using the PAI and WMT contrasts with findings from another sample of postdeployment veterans that found the opposite result: PVTs were unrelated to PTSD and mTBI, but SVT performance (Neurobehavioral Symptom Inventory Validity-10) was related to both conditions (Aase et al., 2020). Conflicting results could reflect different measures used: Aase and colleagues used the several embedded PVTs that are less sensitive than standalone measures. One possibility for these findings in our sample is that the WMT may be more sensitive to globally invalid presentations, or, in contrast, the PAI NIM scale may be relatively insensitive to any garden variety of invalid presentation. To that extent, in the full sample 21.4% had invalid scores on the WMT, whereas only 3.6% had invalid scores on NIM. As a comparison, in Ingram and colleague's (2019) MMPI-2-RF study with veterans, 23.2% were invalid based on F-r, and 12.3% were invalid based on Fp-r. Although lower sensitivity is possibly one explanation, context may explain the difference in base rates, given that the Ingram data was pulled from national clinical records and our sample is predominantly a research-only context. This hypothesis seems to apply to the WMT as well: the Denning and Shura (2019) study found research-context PVT invalidity rates between 5% and 38%, which was on average lower than rates in clinical and forensic samples. Regardless of the results reflecting test or sample characteristics, given the low frequency of invalidating both PVT and SVT indices, results highlight the need for both types of measures in a given evaluation.

Additionally, we evaluated with follow-on analyses the relationship between service-connection status and validity. We established that service-connected disability did not relate to either underperformance on the WMT or overreporting per NIM. Several possible explanations could account for these observations. Most prominently, the context of this evaluation was not for disability, and participants were told such in the consent process, thereby removing the potential link between disability incentive and invalid presentations. Second, service-connected disability encompasses a wide range of physical conditions, and many in the sample who were service connected may have been so only for conditions irrelevant to psychiatric overreporting or cognitive underperformance (e.g., skin conditions). Third, service connection is a complex phenomenon, the nuance of which is not captured in a dichotomous variable.

For example, permanent and total disability status renders a veteran exempt from any future threat of disability benefit reduction. Finally, this variable is based on already-achieved status, as opposed to active claim or intent-to-apply statuses. We evaluated intent-to-apply status, which was also not significant. A study by Critchfield and colleagues (2019) additionally found a non-significant relationship between established disability and invalid performance; however, they also had a variable for active-claim status, which was significant and may be more relevant than our intent-to-apply variable. Future studies might further evaluate these more complex aspects of disability in the VA in relation to validity performance.

Several themes emerge for applying these results. First, the invalid rate on the WMT (≈20%) was consistent with expectations given the known base rates in published research: specifically, veterans and service members in a primarily research context fail standalone PVTs at rates of 5%–38%, with a much lower rate than in primary clinical or forensic samples (Denning & Shura, 2019). The fail rate for the WMT specifically was 39% in the Denning and Shura paper, but that finding was collapsing across contexts. Similarly, invalid rates due to overreporting on the PAI were discrepant from a national sample of veterans using the over-reporting scales on the MMPI-2-RF. In our sample, invalid rates were 3.6% for NIM and 4.3% for MAL; in the Ingram and colleagues (2019) sample, invalid rates were seen at 23.2% for F and 12.3% for Fp. These differential findings might constitute an artifact of context (the Ingram study was derived from clinical data) but may also indicate that the PAI is less sensitive to overreporting than the MMPI-2-RF.

Second, we found a relationship between the WMT and overreporting scales, but not other SVT scales, on the PAI. This finding is not to suggest that the PAI SVT scales can function as overreporting cognitive-complaints scales. For example, Gervais et al. (2007) report on the RBS scale of the MMPI-2/RF, which was designed specifically to detect overreporting cognitive and somatic complaints in disability contexts, and a similar scale has been developed for the PAI (see Gaasedelen et al., 2019). Rather, examinations will yield those who will present in a globally exaggerated fashion involving both self-report and underperformance. Of note, overreporting seems to particularly increase the presence of underperformance: the overall failure rate for the WMT was ≈20%, but in those with invalid PAI profiles, the WMT fail rate rose to 34%. In other words, although very few invalidated both tests, invalidating one does increase the rate of invalidating the other. Thus, especially in cognitive evaluations, if an examinee fails performance validity, this outcome highlights the need to include symptom-validity measures as well. Additionally, the finding that a history of mTBI, particularly when occurring during deployment, was related to WMT

failure, but PTSD symptoms related to both WMT and PAI invalidity, suggests utility in using both types of validity scales within PTSD contexts. Future research might further explore these relationships in other samples.

This study has several limitations. First, our PVT measure was limited to the WMT, a forced-choice memory PVT. Although this WMT has been regarded as the most sensitive PVT by some (Sharland & Gfeller, 2007), it does not capture performance validity in other domains, such as attention and executive functions, and results with other PVTs may be different. Additionally, we relied on a single PVT to identify invalid performance, and therefore some individuals may have yielded false negatives given that invalid performance can wax and wane over time as well as across domains. Thus, we could have missed invalid individuals later in the exam, for example, those rushing to leave later in the day. Although 21.4% of our research-context sample produced invalid scores on the WMT, which is consistent with the expected base rate of failure on a PVT in a veteran/military sample (Denning & Shura, 2019), discrepancies with other studies could follow from our reliance on a single, standalone (though highly sensitive) PVT. Second, although the PAI is established as a well validated multi-scale self-report measure, some consider the MMPI-2/RF as imposing a higher standard given a comparatively greater body of research, and the PAI does not contain well-validated overreporting scales for somatic and cognitive complaints as does the MMPI-2-RF (FBS, RBS), which may be relevant to this population. Furthermore, from a clinical standpoint, one would not interpret ARD-T scores if NIM were invalid; thus, evaluating a relationship in our analyses may be somewhat artificial. Additionally, we neither considered categorical psychiatric diagnoses (including PTSD) nor pain, both of which may be clinically relevant; however, *DSM* diagnostic criteria changed mid-study, and the ARD-T scale is not tied to *DSM-IV-TR* criteria. Additionally, symptom severity and distress levels may be related to both symptom and performance validity more than dichotomous classifications (e.g., Miskey et al., 2020). Finally, TBI status was predominantly determined based on a self-report questionnaire as opposed to more reliable methods such as a structured interview. ¶ Despite these shortcomings, this study adds to the growing body of research on PVTs and SVTs more broadly, and the WMT and PAI more specifically. Clinically, these results highlight that although consideration of both PVTs and SVTs is warranted in most cognitive evaluations, inclusion of both might be particularly important in those reporting PTSD symptoms. Additionally, given the low base rates of failing both PVTs and SVTs, clinicians might be more inclined to pursue either querying about engagement in the evaluation situation or a diagnosis of malingering for patients who present in such a manner.

References

Aase, D. M., Soble, J. R., Shepard, P., Akagi, K., Schroth, C., Greenstein, J. E., Proescher, E., & Phan, K. L. (2020). Concordance of embedded performance and symptom validity tests and associations with mild traumatic brain injury and posttraumatic stress disorder among post-9/11 veterans. *Archives of Clinical Neuropsychology:* 36(3), 424–429. https://doi.org/10.1093/arclin/acaa053

Aikman, G. G., & Souheaver, G. T. (2008). Use of the Personality Assessment Inventory (PAI) in neuropsychological testing of psychiatric outpatients. *Applied Neuropsychology, 15*(3), 176–183. https://doi.org/10.1080/09084280802324283

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

Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society, 11*(3), 215–227. https://doi.org/10.1017/s1355617705050277

Berry, D. T. R., Wetter, M. W., Baer, R. A., Youngjohn, J. R., Gass, C. S., Lamb, D. G., Franzen, M. D., MacInnes, W. D., & Buchholz, D. (1995). Overreporting of closed-head injury symptoms on the MMPI-2. *Psychological Assessment, 7*(4), 517–523. https://doi.org/10.1037/1040-3590.7.4.517

Bowden, S. C., Shores, E. A., & Mathias, J. L. (2006). Does effort suppress cognition after traumatic brain injury? A re-examination of the evidence for the Word Memory Test. *The Clinical Neuropsychologist, 20*(4), 858–872. https://doi.org/10.1080/13854040500246935

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

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

Calhoun, P. S., Earnst, K. S., Tucker, D. D., Kirby, A. C., & Beckham, J. C. (2000). Feigning combat-related posttraumatic stress disorder on the Personality Assessment Inventory. *Journal of Personality Assessment, 75*(2), 338–350. https://doi.org/10.1207/S15327752JPA7502_11

Carone, D. A. (2014). Young child with severe brain volume loss easily passes the Word Memory Test and Medical Symptom Validity Test: Implications for mild TBI. *The Clinical Neuropsychologist, 28*(1), 146–162. https://doi.org/10.1080/13854046.2013.861019

Carroll, L. J., Cassidy, J. D., Peloso, P. M., Borg, J., von Holst, H., Holm, L., Paniak, C., & Pépin, M. (2004). Prognosis for mild traumatic brain injury: Results of the WHO collaborating centre task force on mild traumatic brain injury. *Journal of Rehabilitation Medicine, 36*, 84–105. https://doi.org/10.1080/16501960410023859

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

Critchfield, E., Soble, J. R., Marceaux, J. C., Bain, K. M., Chase Bailey, K., Webber, T. A., et al. (2019). Cognitive impairment does not cause invalid performance: Analyzing performance patterns among cognitively unimpaired, impaired, and noncredible participants across six performance validity tests. *The Clinical Neuropsychologist, 33*(6), 1083–1101. https://doi.org/10.1080/13854046.2018.1508615.

Defense and Veterans Brain Injury Center (DVBIC). (2019). *Total DoD TBI Worldwide Numbers*. Retrieved from https://dvbic.dcoe.mil/dod-worldwide-numbers-tbi.

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

Dismuke-Greer, C. E., Gebregziabher, M., Byers, A. L., Taber, D., Axon, N., Yaffe, K., & Egede, L. E. (2019). Comorbid TBI-depression costs in veterans: A Chronic Effect of Neurotrauma Consortium (CENC) study. *Brain Injury, 33*(2), 198–204. https://doi.org/10.1080/02699052.2018.1542508

Erdodi, L. A., Green, P., Sirianni, C. D., & Abeare, C. A. (2019). The myth of high false-positive rates on the Word Memory Test in mild TBI. *Psychological Injury and Law*. https://doi.org/10.1007/s12207-019-09356-8

Fulton, J. J., Calhoun, P. S., Wagner, H. R., Schry, A. R., Hair, L. P., Feeling, N., Elbogen, E., & Beckham, J. C. (2015). The prevalence of posttraumatic stress disorder in Operation Enduring Freedom/Operation Iraqi Freedom (OEF/OIF) veterans: A meta-analysis. *Journal of Anxiety Disorders, 31*, 98–107. https://doi.org/10.1016/j.janxdis.2015.02.003

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*. https://doi.org/1080/13854046.2019.1612947

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

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

Green, P., & Flaro, L. (2015). Results from three performance validity tests (PVTs) in adults with intellectual deficits. *Applied Neuropsychology: Adult, 22*(4), 293–303. https://doi.org/10.1080/23279095.2014.925903

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

Green, P., Lees-Haley, P. R., & Allen III, L. M. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology, 2*(3–4), 97–124. https://doi.org/10.1300/J151v02n03_05

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology, 18*(2), 86–94. https://doi.org/10.1080/09084282.2010.523389

Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist, 22*(5), 896–918. https://doi.org/10.1080/13854040701565208

Haggerty, K. A., Frazier, T. W., Busch, R. M., & Naugle, R. I. (2007). Relationships among Victoria Symptom Validity Test indices and Personality Assessment Inventory validity scales in a large clinical sample. *The Clinical Neuropsychologist, 21*(6), 917. https://doi.org/10.1080/13854040600899724

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

Hoge, C. W., McGurk, D., Thomas, J. L., Cox, A. L., Engel, C. C., & Castro, C. A. (2008). Mild traumatic brain injury in U.S. Soldiers returning from Iraq. *The New England Journal of Medicine*, *358*(5), 453–463. https://doi.org/10.1056/NEJMoa072972

Hopwood, C. J., Morey, L. C., Rogers, R., & Sewell, K. (2007). Malingering on the Personality Assessment Inventory: Identification of specific feigned disorders. *Journal of Personality Assessment, 88*(1), 43–48. https://doi.org/10.1207/s15327752jpa8801_06

Ingram, P. B., Tarescavage, A. M., 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. *Psychological Services, 17*(3), 355–362. https://doi.org/10.1037/ser0000339

Ivins, B. J., Schwab, K. A., Warden, D., Harvey, S., Hoilien, M., Powell, J., Johnson, E. W., & Salazar, A. M. (2003). Traumatic brain injury in U.S. Army paratroopers: Prevalence and character. *The Journal of Trauma, 55*(4), 617–621. https://doi.org/10.1097/01.TA.0000052368.97573.D4

Jurick, S. M., Crocker, L. D., Keller, A. V., Hoffman, S. N., Bomyea, J., Jacobson, M. W., & Jak, A. J. (2019). The Minnesota Multiphasic Personality Inventory-2-RF in treatment-seeking veterans with history of mild traumatic brain injury. *Archives of Clinical Neuropsychology, 34*(3), 366–380. https://doi.org/10.1093/arclin/acy048

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., & 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

Kucharski, L. T., Toomey, J. P., Fila, K., & Duncan, S. (2007). Detection of malingering of psychiatric disorder with the Personality Assessment Inventory: An investigation of criminal defendants. *Journal of Personality Assessment, 88*(1), 25–32. https://doi.org/10.1207/s15327752jpa8801_04

Lange, R. T., Pancholi, S., Bhagwat, A., Anderson-Barnes, V., & French, L. M. (2012). Influence of poor effort on neuropsychological test performance in U.S. military personnel following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 34*(5), 453–466. https://doi.org/10.1080/13803395.2011.648175

Larrabee, G. J. (2003a). Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *The Clinical Neuropsychologist, 17*(1), 54–68. https://doi.org/10.1076/clin.17.1.54.15627

Larrabee, G. J. (2003b). Exaggerated MMPI-2 symptom report in personal injury litigants with malingered neurocognitive deficit. *Archives of Clinical Neuropsychology, 18*(6), 673–686. https://doi.org/10.1016/S0887-6177(02)00157-9

Lew, H. L., Otis, J. D., Tun, C., Kerns, R. D., Clark, M. E., & Cifu, D. X. (2009). Prevalence of chronic pain, posttraumatic stress disorder, and persistent postconcussive symptoms in OIF/OEF veterans: Polytrauma clinical triad. *Journal of Rehabilitation Research and Development, 46*(6), 697–702. https://doi.org/10.1682/jrrd.2009.01.0006

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

Martindale, S. L., Epstein, E. L., Taber, K. H., VA Mid-Atlantic MIRECC Workgroup, & Rowland, J. A. (2018). Behavioral and health outcomes associated with deployment and nondeployment acquisition of traumatic brain injury in Iraq and Afghanistan veterans. *Archives of Physical Medicine and Rehabilitation, 99*(12), 2485–2495. https://doi.org/10.1016/j.apmr.2018.04.029

Martindale, S. L., Ord, A. S., Lad, S. S., Miskey, H. M., Taber, K. H., & Rowland, J. A. (2020). Differential effects of deployment and nondeployment mild TBI on neuropsychological outcomes. *Rehabilitation Psychology, in Press.* https://doi.org/10.1037/rep0000374

McCaffrey, R. J., O'Bryant, S. E., Ashendorf, L., & Fisher, J. M. (2003). Correlations among the TOMM, Rey-15, and MMPI-2 validity scales in a sample of TBI litigants. *Journal of Forensic Neuropsychology, 3*(3), 45–53. https://doi.org/10.1300/J151v03n03_03

McCormick, C. L., Yoash-Gantz, R. E., McDonald, S. D., Campbell, T. C., & Tupler, L. A. (2013). Performance on the Green Word Memory Test following Operation Enduring Freedom/Operation Iraqi Freedom-era military service: Test failure is related to evaluation context. *Archives of Clinical Neuropsychology, 28*(8), 808–823. https://doi.org/10.1093/arclin/act050

Miskey, H. M., Martindale, S. L., Shura, R. D., & Taber, K. H. (2020). Distress tolerance and symptom severity mediate failure on a symptom validity test in Iraq and Afghanistan veterans with posttraumatic stress disorder. *Journal of Neuropsychiatry and Clinical Neuroscience, 32*, 161-167. https://doi.org/10.1176/appi.neuropsych.17110340

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. https://doi.org/10.1093/milmed/usy038

Morey, L. C. (2007). *The Personality Assessment Inventory professional manual.* Psychological Assessment Resources.

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *The Clinical Neuropsychologist, 17*(2), 263–272. https://doi.org/10.1076/clin.17.2.263.16511

Nelson, N. W., Sweet, J. J., Berry, D. T. R., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society, 13*(3), 440–449. https://doi.org/10.1017/S1355617707070373

Panayiotou, A., Jackson, M., & Crowe, S. F. (2010). A meta-analytic review of the emotional symptoms associated with mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 32*(5), 463–473. https://doi.org/10.1080/13803390903164371

Pugh, M. J., Finley, E. P., Copeland, L. A., Wang, C. P., Noel, P. H., Amuan, M. E., & Pugh, J. A. (2014). Complex comorbidity clusters in OEF/OIF veterans: The polytrauma clinical triad and beyond. *Medical Care, 52*(2), 172–181. https://doi.org/10.1097/mlr.0000000000000059

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*(2), 213–223. https://doi.org/10.1016/j.acn.2006.12.004

Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology, 26*(8), 774–789. https://doi.org/10.1093/arclin/acr066

Sumanti, M., Boone, K. B., Savodnik, I., & Gorsuch, R. (2006). Non-credible psychiatric and cognitive symptoms in a workers' compensation "stress" claim sample. *The Clinical Neuropsychologist, 20*(4), 754–765. https://doi.org/10.1080/13854040500428467

Thomas, K. M., Hopwood, C. J., Orlando, M. J., Weathers, F. W., & McDevitt-Murphy, M. E. (2012). Detecting feigned PTSD using the Personality Assessment Inventory. *Psychological Injury and Law, 5*(3–4), 192–201. https://doi.org/10.1007/s12207-011-9111-6

Van Dyke, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist, 27*(8), 1234–1246. https://doi.org/10.1080/13854046.2013.835447

Veterans Benefits Administration. (2019). *Annual Benefits Report: Fiscal Year 2018*. Retrieved from https://www.benefits.va.gov/REPORTS/abr/docs/2018-abr.pdf

Wang, E. W., Rogers, R., Giles, C. L., Diamond, P. M., Herrington-Wang, L. E., & Taylor, E. R. (1997). A pilot study of the Personality Assessment Inventory (PAI) in corrections: Assessment of malingering, suicide risk, and aggression in male inmates. *Behavioral Sciences and the Law, 15*(4), 469–482. https://doi.org/10.1002/(SICI)1099-0798(199723/09)15:4%3c469::AID-BSL279%3e3.0.CO;2-A

Whiteside, D. M., Dunbar-Mayer, P., & Waters, D. P. (2009). Relationship between TOMM performance and PAI validity scales in a mixed clinical sample. *The Clinical Neuropsychologist, 23*(3), 523–533. https://doi.org/10.1080/13854040802389169

Whiteside, D. M., Galbreath, J., Brown, M., & Turnbull, J. (2012). Differential response patterns on the Personality Assessment Inventory (PAI) in compensation-seeking and non-compensation-seeking mild traumatic brain injury patients. *Journal of Clinical and Experimental Neuropsychology, 34*(2), 172–182. https://doi.org/10.1080/13803395.2011.630648

Youngjohn, J. R., Davis, D., & Wolf, I. (1997). Head injury and the MMPI-2: Paradoxical severity effects and the influence of litigation. *Psychological Assessment, 9*(3), 177–184. https://doi.org/10.1037/1040-3590.9.3.177