Journal of the International Neuropsychological Society (2024), **30**, 410–419
doi:10.1017/S1355617723000711

Research Article

Importance of validity testing in psychiatric assessment: evidence from a sample of multimorbid post-9/11 veterans

Sahra Kim[1], Alyssa Currao[1], Emma Brown[1], William P. Milberg[1,2] and Catherine B. Fortier[1,2]

[1]Translational Research Center for TBI and Stress Disorders and Geriatric Research Education and Clinical Center, VA Boston Healthcare System, Boston, MA, USA and [2]Department of Psychiatry, Harvard Medical School, Boston, MA, USA

Abstract

Objective: Performance validity (PVTs) and symptom validity tests (SVTs) are necessary components of neuropsychological testing to identify suboptimal performances and response bias that may impact diagnosis and treatment. The current study examined the clinical and functional characteristics of veterans who failed PVTs and the relationship between PVT and SVT failures. **Method:** Five hundred and sixteen post-9/11 veterans participated in clinical interviews, neuropsychological testing, and several validity measures. **Results:** Veterans who failed 2+ PVTs performed significantly worse than veterans who failed one PVT in verbal memory (Cohen's $d = .60–.69$), processing speed (Cohen's $d = .68$), working memory (Cohen's $d = .98$), and visual memory (Cohen's $d = .88–1.10$). Individuals with 2+ PVT failures had greater posttraumatic stress (PTS; $\beta = 0.16$; $p = .0002$), and worse self-reported depression ($\beta = 0.17$; $p = .0001$), anxiety ($\beta = 0.15$; $p = .0007$), sleep ($\beta = 0.10$; $p = .0233$), and functional outcomes ($\beta = 0.15$; $p = .0009$) compared to veterans who passed PVTs. 7.8% veterans failed the SVT (Validity-10; ≥19 cutoff); Multiple PVT failures were significantly associated with Validity-10 failure at the ≥19 and ≥23 cutoffs (p's < .0012). The Validity-10 had moderate correspondence in predicting 2+ PVTs failures ($AUC = 0.83$; *95% CI* = 0.76, 0.91). **Conclusion:** PVT failures are associated with psychiatric factors, but not traumatic brain injury (TBI). PVT failures predict SVT failure and vice versa. Standard care should include SVTs and PVTs in all clinical assessments, not just neuropsychological assessments, particularly in clinically complex populations.

Keywords: Veteran health; neuropsychological tests; performance validity tests; symptom validity tests; post-traumatic stress disorder; psychiatric disorders

(Received 1 May 2023; final revision 26 September 2023; accepted 9 October 2023; First Published online 28 November 2023)

From 2000 to 2021, approximately 450,000 US service members were diagnosed with a traumatic brain injury (TBI) with the majority (82%) being mild in severity (Traumatic Brain Injury Center of Excellence (TBICoE), 2021). The prevalence of mild traumatic brain injury (mTBI), also known as concussion, in recent era, veterans has led to an increased need for assessments, including neuropsychological evaluations for diagnostic conclusions, which in part determine the distribution of disability benefits, service connection, and access to health care. As of 2015, approximately 100,000 veterans were receiving VA disability compensation for TBIs (Denning & Shura, 2017).

Performance validity tests (PVTs) are a crucial component of neuropsychological evaluations as they measure credible or valid performance and ensure that the results of testing are a true representation of cognitive functioning (Sweet et al., 2021). In addition to the recommendations of the American Academy of Clinical Neuropsychology (Sweet et al., 2021), the Military Traumatic Brain Injury Task Force has recommended the inclusion of validity measures in neuropsychological evaluations given possible external motivation or incentives that may impact the assessment and recovery processes (McCrea et al., 2008). Failing PVTs suggests atypical patterns of test performance that are likely noncredible; in other words, interpreting the neuropsychological testing results may lead to a misdiagnosis. For example, an individual may incorrectly be diagnosed with a neurocognitive disorder. Serious adverse consequences from misdiagnosis may include individuals being referred to inappropriate and costly treatments, depleting healthcare resources, and creating financial burden (Denning & Shura, 2017). A misdiagnosis can also cause significant emotional distress for individuals and their families as well as lead to the unnecessary restriction of independent activities of daily living. Additionally, a misdiagnosis can lead to iatrogenic effects, erroneously reinforcing symptoms that would otherwise not be present, and exacerbating functional decline. Therefore, PVTs are an imperative component in neuropsychological assessment, including TBI assessment.

Invalid performances, or PVT failure, among post-9/11 veterans and service members has ranged widely from 6% to 68% across studies (Armistead-Jehle & Hansen, 2011; Armistead-Jehle, 2010; McCormick et al., 2013; Russo, 2012). Studies in those

Corresponding author: Sahra Kim; Email: Sahra.Kim@va.gov

Cite this article: Kim S., Currao A., Brown E., Milberg W.P., & Fortier C.B. (2024) Importance of validity testing in psychiatric assessment: evidence from a sample of multimorbid post-9/11 veterans. *Journal of the International Neuropsychological Society*, **30**: 410–419, https://doi.org/10.1017/S1355617723000711

with TBI have demonstrated that poor performance on validity tests accounts for much of the variability in neuropsychological testing (Green et al., 2001; Meyers et al., 2011). One study showed that patients with an active compensation claim (72%) demonstrated poorer performance validity compared to those without an active claim (15%; Critchfield et al., 2019). A recent study suggested that applying for disability benefits, which is associated with the motivation for secondary gain, can impact performance validity (Horner et al., 2022). Alternatively, when assessments are completed outside of clinical setting where there are no potential external incentives or financial compensation (e.g., in a research context), PVT failures among post-9/11 veterans are much lower, ranging from 4% to 9% (Clark et al., 2014).

Although suboptimal performance validity can be due to external incentives such as those seeking disability benefits (e.g., increase of service connection), poor performance validity does not equate to malingering and may also be associated with internal, psychiatric factors. For example, among the more than half (58%) of veterans who were positive on TBI screening and performed below the cutoffs on the Medical Symptom Validity Test (MSVT), approximately 69% had depression (Armistead-Jehle, 2010). Another recent study demonstrated that severity of posttraumatic stress (PTS; formerly posttraumatic stress disorder; PTSD) symptoms was associated with MSVT failure (Miskey et al., 2020). Veterans who failed the Word Memory Test (WMT), a verbal memory task similar to the MSVT, had greater prevalence of current PTS and Major Depressive Disorder compared to those who passed (Shura et al., 2016). Furthermore, those with comorbid psychiatric diagnoses (e.g., TBI, PTS, depression) have increased rates of negative response bias (Lange et al., 2012).

Young et al. (2016) reported that 45% of psychologists from the VA (Veterans Affairs) Healthcare System determined that failing even one PVT was sufficient to deem a performance invalid, while 47% used at least two PVT failures as a minimum benchmark. One study examining veterans with mTBI found that there were significant differences on tests of verbal memory, processing speed, and cognitive flexibility among those who passed versus those who failed one PVT (WMT). However, those who failed one PVT compared to two PVTs only differed in one measure of delayed free recall, suggesting that clinicians should consider a performance invalid if individuals failed even a single PVT (Proto et al., 2014). However, several other studies have suggested that failure of two or more PVTs has high specificity and the use of several PVTs increases sensitivity without compromising specificity (Martin et al., 2015; Schroeder & Marshall, 2011). Therefore, noting a failure in two or more independent (e.g., no two embedded PVTs from the same measure) well-validated PVTs is the recommended threshold for detecting invalid cognitive performances (Jennette et al., 2022), as relying on a single PVT may result in high false positive rates (Victor et al., 2009).

Whereas PVTs evaluate the validity of objective cognitive abilities, symptom validity tests (SVTs) evaluate the credibility of subjective reports. Symptom validity tests are used to identify symptom exaggeration or overreporting in self-report measures and should also be regularly utilized in neuropsychological assessments (Boe & Evald, 2022; Larrabee, 2012). However, the use of SVTs is not consistently and routinely used in conjunction with clinical assessments such as the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5) or the Structured Clinical Interview for DSM-V (SCID-5). One study utilizing the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) found that approximately 5–27% of a veteran sample failed validity scales that detect overreporting (Ingram et al., 2020), highlighting the need to include SVTs in all clinical assessments and not limit their use to the field of neuropsychology.

The Neurobehavioral Symptom Inventory (NSI), which assesses self-report of postconcussive symptoms, has been widely used by the DoD and VA in TBI evaluations. The Validity-10 is the most recommended and effective scale within the NSI to detect noncredible reporting of symptoms (Ashendorf, 2019; Lange et al., 2015; Vanderploeg et al., 2014). Symptom validity tests and PVTs are related such that those who perform suboptimally on cognitive testing are more likely to express greater subjective complaints, however, they measure independent constructs (Boe & Evald, 2022; Clark et al., 2014; Ord et al., 2021). Aase et al. (2021) examined performances on four embedded validity measures and their relationship with the Validity-10 in a sample of post-9/11 veterans. Veterans who passed PVTs were more likely to pass the Validity-10 (at ≥13 and ≥19 cutoffs), while veterans who failed at least one embedded PVT were more likely to fail the Validity-10. Additionally, veterans who had both PTS and mTBI were more likely to fail the Validity-10.

The current study first examines cognitive performance based on PVT failure to determine whether there are significant differences in failing one versus two PVTs among a research sample of post-9/11 veterans. Second, the clinical characteristics and functional outcomes of those who failed 2+ PVTs (standalone and embedded measures) are examined within this population. Last, we examine if PVT failure is associated with SVT (NSI; Validity-10) failure using three distinct cutoffs (Lange et al., 2015), and whether SVT failure predicts PVT failure.

Method

Participants

Participants included 813 veterans and National Guard/Reservists who deployed to post-9/11 conflicts (Operations Enduring Freedom, Iraqi Freedom, and New Dawn; this sample will be collectively labeled as "veterans" for simplicity) who were enrolled in Translational Research Center for Traumatic Brain Injury and Stress Disorders (TRACTS) longitudinal cohort study. Participants were recruited primarily from Boston, Massachusetts (New England area) and Houston, Texas by a recruitment specialist who attended military events augmented by the distribution of flyers within the VA Healthcare Systems and the greater community (for more details, please see McGlinchey et al., 2017). The sample includes veterans from over 30 U.S. states and is reflective of post-9/11 era military demographics. Veterans were excluded for a history of neurological disorder (with the exception of TBI), seizure disorder (not related to TBI), significant psychiatric conditions (e.g., bipolar disorder, psychotic disorders), or active suicidal or homicidal ideations. Participants are from a research sample where primary and secondary gain has been minimized; they were informed that research evaluations were not documented in clinical medical records and therefore had no impact on establishing or increasing disability benefits. This study has been approved by the VA Boston Institutional Review Board for human participants' protection. All study procedures were completed in accordance with the Declaration of Helinski principles.

For the present study, we removed participants who were only administered a limited set of PVTs (MSVT, CVLT-II) at the Houston assessment site ($n = 177$). We further excluded participants with a moderate or severe TBI (e.g., loss of consciousness

https://doi.org/10.1017/S1355617723000711 Published online by Cambridge University Press

>30 minutes, alteration of mental status >24 hours, posttraumatic amnesia >24 hours; $n = 26$), non-native English speakers ($n = 2$), and participants who had a personality disorder or other significant psychiatric concern ($n = 4$), neurologic condition (e.g., heavy metal exposure, brain atrophy evident in imaging scan; $n = 3$), or concerns related to the accuracy of the clinical interview ($n = 1$). An additional 84 participants did not complete PVTs (e.g., MSVT, CVLT-II, BVMT-R, Digit Span) due to time constraints and were therefore excluded from the current analysis, yielding a final sample size of 516.

Measures

Psychological assessments

The diagnoses of PTS, TBI, and other psychiatric conditions were assessed via clinical interviews administered by a doctoral-level clinician. The Clinician-Administered PTSD Scale for DSM-IV (CAPS-IV) assessed for PTS (Blake et al., 1995), the Boston Assessment of Traumatic Brain Injury-Lifetime (BAT-L) assessed history of TBI (Fortier et al., 2014), and the Structured Clinical Interview for DSM-IV/V (SCID-IV/V; First et al., 1997) assessed mental health disorders including mood and anxiety disorders. Clinical interviews at both sites were reviewed in diagnostic consensus meetings with at least three doctoral-level clinicians.

Neuropsychological testing

Participants in TRACTS were administered a fixed neuropsychological battery measuring the cognitive domains of verbal (California Verbal Learning Test – Second Edition; CVLT-II; Delis et al., 2000) and visual memory (Brief Visuospatial Memory Test – Revised; BVMT-R; Benedict, 1997), attention/working memory (e.g., digit span and coding from the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008)), executive functioning (e.g., verbal fluencies including letter, category, and category switching and trail making tests including number sequencing and number letter sequencing from Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001)), Grooved Pegboard (Tiffen, 1968), and Auditory Consonant Trigram (ACT; Stuss et al., 1985). The Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) was administered to provide a measure of premorbid functioning.

Performance validity tests

Participants were given a stand-alone, computer-administered PVT, the Medical Symptom Validity Test (MSVT), which evaluated level of test engagement (Green, 2004). Cutoffs suggesting suboptimal performances on the MSVT are described in the manual. All neuropsychological tests and PVTs in the standard battery were administered in the same order to all participants.

Among the embedded PVTs, a systematic review informed a cutoff score of ≤14 (sensitivity 50% and specificity 93%) on the CVLT-II Forced choice (Schwartz et al., 2016). In the BVMT-R, a cutoff score of ≤4 in the recognition discrimination index (sensitivity 50% and specificity 93%) or ≤4 recognition hits (sensitivity 45% and specificity 89%) identified noncredible performances (Bailey et al., 2018; Denning, 2012). A retention rate of ≤58% in the BVMT-R (sensitivity 31% and specificity 92%) was also identified as a cutoff for embedded PVT failure (Sawyer et al., 2017). Lastly, a cutoff score of ≤6 (sensitivity 54% and specificity 91%) on the reliable digit span (RDS) from the WAIS-IV Digit Span, which measures attention and working memory, was identified as a PVT failure (Webber & Soble, 2018; Wechsler, 2008).

Symptom validity test

The Validity-10 from the NSI includes unlikely and low-frequency items (e.g., items that are uncommonly endorsed) that can identify symptom exaggeration; failure of the Validity-10 may prompt further follow up (Lange et al., 2015; Vanderploeg et al., 2014). Lange and colleagues (2015) suggested that a cutoff score of ≥19 indicated "possible exaggeration" (59% sensitivity; 89% specificity; 74% positive predictive value (PPV); 80% negative predictive value (NPV)), ≥23 indicated "probable exaggeration" (41% sensitivity; 96% specificity; 75% PPV; 83% NPV), and ≥28 indicated "highly probable exaggeration" (22% sensitivity; 99% specificity; 94% PPV; 70% NPV).

Self-report questionnaires

Self-report questionnaires included the Depression Anxiety Stress Scale-21 (DASS-21; Henry & Crawford, 2005), Lifetime Drinking History (LDH; Skinner & Sheu, 1982), McGill Pain Questionnaire (short form; Melzack, 1975), Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989), Neurobehavioral Symptom Inventory (NSI; Cicerone, 1995), and the WHO Disability Assessment Scale-II (WHODAS-II; Üstün et al., 2010).

Statistical analyses

To compare PVT cutoffs, independent t-tests and effect sizes (Cohen's *d*) were used to examine differences in neuropsychological performance for the following pairwise combinations of PVT groups: (1) no failed PVTs vs. failed 1 PVT, (2) no failed PVTs vs. failed 2+ PVTs, (3) failed 1 vs. 2+ PVTs (similar to Proto et al. (2014)). Cohen's *d* for unequal variance was calculated when comparison groups did not meet equal variance assumptions. Analyses were conducted on norm-standardized scores. Since the RDS score was derived from Digit Span, it was not included as part of the neuropsychological variables (Table 2). (However, the CVLT-II was included because CVLT-II Forced Choice is a separate trial within the CVLT-II and not directly derived from the total recall and long delay trials. Similarly, the embedded measures from the BVMT-R are not derived from the total or delayed recall.) Additionally, we calculated the area under the curve and 95% confidence intervals using logistic regression models to evaluate the use of the Validity-10 to predict failure for 1+ and 2+ PVTs.

Differences in demographic and clinical characteristics between PVT groups (e.g., passed vs. failed 2+ PVTs) were determined using independent t-tests for continuous variables and chi-square for categorical variables. Fisher's exact test was used for categorical variables when an expected cell count was less than 5. Similar to Clark et al. (2014), we used linear regression models to examine differences in psychological symptom severity, somatic, and functional outcomes after controlling for age and education. For outcomes that did not meet linear regression assumptions, we applied a square root transformation to normalize the residuals. As a sensitivity analysis, we examined whether differences in outcomes persisted after removing SVT failures using all three cutoffs. Additionally, we explored whether standalone or embedded performance validity measures better-predicted differences in outcomes. All p-values refer to two-tailed tests. Statistical analyses were conducted in SAS (version 9.4)

https://doi.org/10.1017/S1355617723000711 Published online by Cambridge University Press

Table 1. Demographics

Covariates	*n*	Full sample Mean (*SD*)
Education (years)	516	14.1 (2.1)
WTAR standard score	515	104.2 (11.8)
% service connection	312	44.2 (36.0)
n Post-9/11 deployments	516	1.6 (1.0)
Times since last post-9/11 deployment (months)	495	49.8 (42.6)
		***N* (%)**
Male	516	458 (88.8%)
Ethnicity		
Hispanic/Latino	516	83 (16.1%)
Race		
White	516	390 (75.6%)
Black/African American	516	50 (9.7%)
Asian	516	13 (2.5%)
American Indian/Alaskan	516	6 (1.2%)
Hawaiian/Pacific Islander	516	3 (0.6%)
Other	142	7 (4.9%)
Military branch		
Army	516	344 (66.7%)
Navy	516	26 (5.0%)
Air force	516	49 (9.5%)
Marines	516	105 (20.4%)
Coast guard	516	1 (0.2%)
National guard/reserves	516	255 (49.4%)

Note. SD = Standard Deviation; WTAR = Wechsler Test of Adult Reading.

Results

Participants were largely male (88.8%) and white (75.6%) and representative of U.S. military demographics. The average education was 14.1 years (Standard Deviation [*SD*] = 2.1), and estimated premorbid intelligence measured by the Weschler Test of Adult Reading (WTAR; Wechsler, 2001) was 104.2 (*SD* = 11.8). Demographic information is presented in Table 1.

Participants who failed one PVT test performed significantly worse than those who failed none on the CVLT-II total trials and long delay free recall; DKEFS letter fluency, category fluency, category switching, number sequencing, and number/letter switching; WAIS-IV coding; ACT total score on 0–36 s delay; Grooved Pegboard dominant hand trial; and BVMT-R total recall and delayed recall. Effect sizes for these differences ranged from small to medium (Cohen's $d = 0.28–0.62$; Cohen, 1988). Similarly, participants who failed 2+ PVTs performed significantly worse on all neuropsychological measures except for the Grooved Pegboard compared to counterparts who failed none. Effect sizes were larger for the 2+ PVT failure group, with Cohen's *d* estimates ranging from 0.82 to 2.02. Participants who failed two or more PVTs performed significantly worse than those who failed one PVT on all measures except Grooved Pegboard and DKEFS letter fluency and number/letter switching. These effect sizes ranged from medium to large (Cohen's $d = 0.60–1.10$; see Table 2).

Among the 516 participants, 5.4% ($n = 28$) of participants failed the RDS, 4.8% ($n = 25$) failed the MSVT, 4.8% ($n = 25$) failed the BVMT-R recognition discrimination index, and 2.1% ($n = 11$) failed the CVLT-II forced choice, 1.9% ($n = 10$) failed the BVMT-R recognition hits, and 1.4% ($n = 7$) failed the BVMT-R percent retention. Veterans who failed 2+ PVTs ($n = 17$) had less education (*Mean* = 12.9 years vs. 14.2 years; $p = .0114$) and lower WTAR standard scores (*Mean* = 97.6 vs. 104.4; $p = .0183$; see Table 3). They also differed in clinical characteristics such that those with multiple PVT failures were more likely to have PTS diagnoses (88.2% vs. 55.4%; $p = .0073$) as well as greater PTS severity (*Mean* = 77.7 vs. 47.3; $p < .0001$), mood disorders (64.7% vs. 25.1%; $p = .0008$), and deployment trauma phenotype (DTP; also known as comorbid depression, PTS, and military-related mTBI diagnoses; 35.3% vs. 14.6%; $p = .0327$). Participants who failed 2+ PVTs also reported greater pain (*Mean* = 51.9 vs. 30.8; $p = .0012$), sleep disturbances (*Mean* = 13.6 vs. 9.8; $p = .0036$), and functional impairment (*Mean* = 36.9 vs. 17.9; $p < .0001$). Notably, there were no differences in the prevalence of lifetime or miliary-related mTBI based on PVT failure. After adjusting for age and education, CAPS-IV PTS symptom severity ($\beta = 0.16$; $p = .0002$) and self-reported depression ($\beta = 0.17$; $p = .0001$) and anxiety symptoms ($\beta = 0.15$; $p = .0007$) were higher among those who failed 2+ PVTs (see Table 4). Furthermore, they had greater sleep disturbances ($\beta = 0.10$; $p = .0233$) and worse functional impairment ($\beta = 0.15$; $p = .0009$).

Among a subset of 488 participants who completed the SVT, 7.8% ($n = 38$) failed using a Validity-10 cutoff score of ≥19, 3.3% ($n = 16$) failed using a cutoff score of ≥23, and 1.4% ($n = 7$) failed using a cutoff score of ≥28. Multiple PVT failures were significantly associated with Validity-10 failure when using the ≥19 and ≥23 cutoffs (*p*'s < .0012), but not the ≥28 cutoff. Additionally, we looked at the area under the curve (AUC) to evaluate how well the Validity-10 predicted PVT failures. AUC values greater than 0.9 indicate high discrimination, values between 0.7 and 0.9 indicate moderate discrimination, and values below 0.7 indicate poor discrimination between measures (Fischer et al., 2003; Swets, 1988). The Validity-10 had poor correspondence with failing one or more PVTs (*AUC* = 0.65; *95% Confidence Interval [CI]* = 0.58, 0.73). However, the Validity-10 had moderate correspondence with failing two or more PVTs (*AUC* = 0.83; *95% CI* = 0.76, 0.91).

We conducted a sensitivity analysis by removing Validity-10 failures using all three cutoffs (≥19, ≥23, and ≥28). Once the Validity-10 failures were removed, we examined the association between multiple PVT failures and clinical characteristics to see if any associations changed. After removing Validity-10 scores ≥19, failing 2+ PVTs was associated with higher PTS symptom severity ($\beta = 0.14$; $p = .0036$), self-reported depression symptoms ($\beta = 0.12$; $p = .0107$), and functional impairment ($\beta = 0.10$; $p = .0361$). However, self-reported anxiety symptoms and sleep disturbances were no longer significant. After removing Validity-10 scores ≥23, PTS symptom severity ($\beta = 0.13$; $p = .0041$), self-reported depression ($\beta = 0.14$; $p = .0027$) and anxiety symptoms ($\beta = 0.10$; $p = .0407$), and functional impairment ($\beta = 0.12$; $p = .0072$) were higher among those with multiple failures, but sleep disturbances were no longer associated with multiple PVT failures. Finally, after Validity-10 scores ≥28 were removed, 2+ PVT failure was associated with higher PTS symptom severity ($\beta = 0.15$; $p = .0008$), self-reported depression ($\beta = 0.16$; $p = .0005$) and anxiety symptoms ($\beta = 0.14$; $p = .0031$), sleep disturbances ($\beta = 0.10$; $p = .0363$), and functional impairment ($\beta = 0.14$; $p = .0017$).

We also further we examined the association between failing one or more measure on the standalone MSVT measure versus an embedded measure within the WAIS-IV (RDS), CVLT-II, or BVMT-R. PTS symptom severity and self-reported depression and anxiety were higher among participants regardless of whether they failed the standalone MSVT or one of the embedded measures (*p*'s < .02). Any failure was associated with greater pain severity and worse sleep disturbances and functional impairment for both standalone and embedded measures (*p*'s < .02). For all psychiatric, somatic, and functioning outcomes, failure on the standalone

https://doi.org/10.1017/S1355617723000711 Published online by Cambridge University Press

Table 2. Descriptive and effect sizes for pairwise combinations of performance validity test (PVT) failure groups

PVT failure group	CVLTTL(T)	CVLTLD(Z)	FAS LF(SS)	FASCF(SS)	FASCS(SS)	TMTNS(SS)	TMTNLS (SS)	ACT	GRV DH (T)	GRV NDH (T)	BVMTTR(T)	BVMT DR(T)
N	516	516	510	510	509	507	507	355	514	513	516	516
0	49.5 (10.2)	−0.2 (1.1)	11.2 (3.5)	12.0(3.3)	11.0(3.2)	11.2 (2.3)	10.3 (2.6)	43.3 (8.6)	44.4 (10.9)	45.0 (11.4)	48.4 (11.4)	52.5 (10.5)
1	45.1 (9.7)	−0.8 (1.2)	9.4 (3.4)	10.5(4.2)	9.8(2.9)	10.0 (2.4)	8.8 (2.6)	38.0 (8.0)	41.4 (9.4)	43.9 (12.1)	43.8 (13.6)	46.6 (14.2)
2+	39.6 (7.2)	−1.6 (1.1)	7.8 (3.2)	7.8 (3.1)	7.8 (3.4)	7.9 (3.8)	7.6 (3.8)	30.2 (7.5)	42.4 (14.2)	42.9 (11.7)	32.1 (12.5)	31.1 (13.5)
0 vs. 1	.427*	.542*	.510*	.395*	.390*	.515*	.572*	.619*	.283*	.092	.396*	.472*
0 vs. 2+	.970*	1.314*	.954*	1.278*	.995*	*1.069**	*.815**	1.523*	.187	.180	1.427*	2.019*
1 vs. 2+	.595*	.689*	.466	.692*	.647*	.676*	.359	.984*	*.082*	.083	.875*	1.104*

Note. CVLT TL = California Verbal Learning Test – Second Edition (CVLT-II) trials 1–5 total learning; CVLT LD = CVLT-II long delay free recall; FAS LF = letter fluency; FAS CF = category fluency; FAS CS = category switching; TMT NS = trails making test number sequencing total time; TMT NLS = trails making test number/letter switching total time; DS = Wechsler Adult Intelligence Scale – Fourth Edition Coding; ACT = Auditory Consonant Trigram 0–36 s delay total; GRV DH = Grooved Pegboard dominant hand total time; GRV NDH = Grooved Pegboard non-dominant hand total time; BVMT TR = Brief Visuospatial Memory Test – Revised (BVMT-R) total recall; BVMT TR = BVMT-R delayed recall. T indicates *t*-score. Z indicates *z*-score. SS indicates standard score. Italics = Cohen's *d* for unequal variance.
*$p < .05$

MSVT was associated with a greater increase in impairment scores as compared to an embedded measure.

Discussion

With the high prevalence of head injuries sustained during post-9/11 conflicts, there is a demand for TBI assessment including neuropsychological evaluations. PVTs are necessary components of TBI assessment as they can detect suboptimal performances affecting the interpretation of the test data and ultimately clinical decision making and service connection status (Sweet et al., 2021). Approximately 15% of veterans with a TBI failed at least one PVT as did 10% of veterans without TBI. TBI was not associated with failing 2+ PVTs, further suggesting that history of TBI did not play a significant role in PVT failure in our sample. Our findings were similar to previous studies showing that PVT failure rates were much lower in a research setting (ranging from 1.4% to 5.4% failure rates in any one of the PVTs administered) compared to forensic or clinical settings where medical records may be used to determine disability compensation (Clark et al., 2014; Denning & Shura, 2017; McCormick et al., 2013). Only 17 veterans in the research sample failed 2+ PVTs; due to the low rate of failures, there are limits to generalizability in other study populations as well as clinical veteran populations where there may be motivation for secondary gain. It remains unclear what proportion of participants believed that there were no potential external incentives as a participant in research.

Proto et al. (2014) suggested that failing even one PVT, specifically the WMT could invalidate neuropsychological results, however, our findings strengthen the recommendation of using a threshold of 2+ PVT failures for detecting noncredible cognitive performances in a veteran research sample. This study examined the incidence of failure across PVTs from four different tests. Effect sizes were larger when comparing the no PVT failure group to the 2+ PVT failure group. Among our veteran research sample, those who failed multiple PVTs performed worse on most cognitive measures compared to those who failed one PVT, with medium to large effect sizes, suggesting that the cutoff of 2+ PVTs should be used to determine assessment invalidity in this population. Since performance and testing engagement may change over time and throughout the evaluation (Boone, 2009), clinicians are recommended to utilize multiple PVT (both standalone and embedded; Critchfield et al., 2019; Sweet et al., 2021), across various neuropsychological domains. Clinicians are also recommended to use the appropriate cutoffs considering the sensitivity and specificity (as well as positive and negative predictive value) of measures in a given population (e.g., intellectual disability, mild cognitive impairment or dementia (Dean et al., 2009), English as a second language (Lippa, 2018)). PVTs are designed to have greater specificity (at least 90%) compared to sensitivity as it minimizes the number of false positives to avoid erroneously labeling someone as potentially malingering. In our sample, failing 2+ PVTs increases certainty that performances in cognitive testing is invalid and should not be interpreted as results likely underestimate true ability (Boone, 2021). Providers who only use a single PVT failure as a minimum criterion may be overclassifying test performances as invalid (Young et al., 2016).

Post-9/11 veterans who failed 2+ PVTs had significantly higher rates of PTS as well as greater severity of PTS symptoms (e.g., higher CAPS-IV scores) and diagnosable mood disorders with higher self-reported depression and anxiety symptoms

https://doi.org/10.1017/S1355617723000711 Published online by Cambridge University Press

Table 3. Demographics and clinical characteristics stratified by participants who failed two or more performance validity tests (PVTs)*

		Passed ($n = 499$)	Failed 2+ PVTs ($n = 17$)	
Covariates	*n*	Mean *(SD)*	Mean *(SD)*	*P*-value
Education (years)	516	14.2 (2.1)	12.9 (1.6)	.0114
WTAR Standard Score	515	104.4 (11.6)	97.6 (15.3)	.0183
% Service Connection	312	43.5 (35.8)	66.7 (36.4)	.0573
Psychiatric				
CAPS-IV PTS Total	514	47.3 (29.8)	77.7 (17.5)	<.0001
DASS-21 Depression	494	8.6 (9.6)	20.4 (11.6)	<.0001
DASS-21 Anxiety	494	6.7 (7.6)	15.6 (11.4)	.0072
LDH (weight-corrected)	509	2125.9 (3661.8)	2591.6 (2745.8)	.6039
Somatic				
McGill Pain Severity	478	30.8 (25.4)	51.9 (30.7)	.0012
PSQI Sleep Disturbance	489	9.8 (4.8)	13.6 (4.1)	.0036
Functioning (WHODAS)				
Subjective Health Rating	461	2.3 (0.8)	2.9 (0.9)	.0145
Functional Impairment	488	17.9 (16.3)	36.9 (18.4)	<.0001
		N (%)	*N* (%)	
Male	516	441 (88.4)	17 (100.0)	.2394
Ethnicity				
Hispanic/Latino	516	80 (16.0)	3 (17.7)	.7442
Race				
White	516	377 (75.6)	13 (76.5)	.9999
Black/African American	516	49 (9.8)	1 (5.9)	.9999
Asian	516	13 (2.6)	0 (0.0)	.9999
American Indian/Alaskan	516	6 (1.2)	0 (0.0)	.9999
Hawaiian/Pacific Islander	516	3 (0.6)	0 (0.0)	.9999
Other	142	7 (5.1)	0 (0.0)	.9999
mTBI				
Lifetime	516	336 (67.3)	14 (82.4)	.1924
Military	516	216 (43.3)	11 (64.7)	.0802
Current diagnoses				
CAPS-IV PTS	515	276 (55.4)	15 (88.2)	.0073
SCID Mood Disorder	515	125 (25.1)	11 (64.7)	.0008
SCID Anxiety Disorder	515	86 (17.3)	5 (29.4)	.1990
SCID Alcohol Use Disorder	515	64 (12.9)	5 (29.4)	.0634
SCID Non-Alcohol Use Disorder	515	17 (3.4)	0 (0.0)	.9999
Deployment Trauma Phenotype (Current PTS, Current Depression, and Military mTBI)	495	70 (14.6)	6 (35.3)	.0327

Note. PVT = performance validity test; WTAR = Wechsler Test of Adult Reading; CAPS-IV = Clinician-Administered PTSD Scale for DSM-IV; PTS = posttraumatic stress; DASS-21 = Depression, Anxiety, and Stress Scale – 21 items; LDH = Lifetime Drinking History; PSQI = Pittsburgh Sleep Quality Index; WHODAS = World Health Organization Disability Assessment Scale; mTBI = mild traumatic brain injury; SCID = Structured Clinical Interview for DSM-IV.
*A PVT failure was considered a (1) Medical Symptom Validity Test (MSVT) immediate recognition, delayed recognition, or consistency index ≤85%; or a (2) Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV) reliable digit span score ≤6; or a (3) California Verbal Learning Test – Second Edition (CVLT-II) forced choice score ≤14; or a (4) Brief Visuospatial Memory Test – Revised (BVMT-R) recognition discrimination index score ≤4, recognition hits score ≤4, or percent retained ≤58%.

Table 4. Adjusted linear regression analyses for 2+ performance validity test failure

		Failed 2+ PVTs*				
Covariates	*n*	R^2	*B*	*SE*	β	*P*-value
Psychiatric						
CAPS-IV PTS Total	514	0.07	27.03	7.33	0.16	.0002
DASS-21 Depression	494	0.05	1.77	0.46	0.17	.0001
DASS-21 Anxiety	494	0.04	1.44	0.42	0.15	.0007
Somatic						
McGill Pain Severity	478	0.03	1.38	0.74	0.09	.0624
PSQI Sleep Disturbance	489	0.05	2.95	1.30	0.10	.0233
WHODAS Functional Impairment	488	0.06	1.77	0.53	0.15	.0009

Note. PVT = performance validity test; SE = standard error; CAPS-IV = Clinician-Administered PTSD Scale for DSM-IV; PTS = posttraumatic stress; DASS-21 = Depression, Anxiety, and Stress Scale – 21 items; PSQI = Pittsburgh Sleep Quality Index; WHODAS = World Health Organization Disability Assessment Scale.
*A PVT failure was considered a (1) Medical Symptom Validity Test (MSVT) immediate recognition, delayed recognition, or consistency index ≤85%; or a (2) Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV) reliable digit span score ≤6; or a (3) California Verbal Learning Test – Second Edition (CVLT-II) forced choice score ≤14; or a (4) Brief Visuospatial Memory Test – Revised (BVMT-R) recognition discrimination index score ≤4, recognition hits score ≤4, or percent retained ≤58%.
Models are adjusted for age and education.

(based on self-report questionnaires) compared to those passed. Greater physical pain, poorer sleep quality, and lower overall functional outcomes were also significantly associated with 2+ PVT failures. Results are consistent with several prior studies highlighting the link between poor PVT performances and clinical psychiatric factors (Armistead-Jehle, 2010; Miskey et al., 2020). Furthermore, having multiple PVT failures were also associated with a trio of diagnoses consisting of PTS, mTBI, and mood

https://doi.org/10.1017/S1355617723000711 Published online by Cambridge University Press

Table 5. Adjusted standardized betas for failure on standalone and embedded measures

		Any failure							
		Standalone (MSVT)				Embedded (WAIS-IV, CVLT-II, or BVMT-R)			
Covariates	*n*	*B*	*SE*	*β*	*P*-value	*B*	*SE*	*β*	*P*-value
Psychiatric									
CAPS-IV PTS Total	514	19.50	6.01	0.14	.0012	12.52	4.07	0.13	.0022
DASS-21 Depression	494	1.55	0.37	0.18	<.0001	0.64	0.26	0.11	.0136
DASS-21 Anxiety	494	1.15	0.34	0.15	.0008	0.66	0.24	0.13	.0051
Somatic									
McGill Pain Severity	478	1.69	0.61	0.13	.0057	1.34	0.41	0.15	.0011
PSQI Sleep Disturbance	489	2.47	1.01	0.11	.0146	1.85	0.67	0.12	.0060
WHODAS Functional Impairment	488	1.95	0.43	0.20	<.0001	1.02	0.29	0.16	.0005

Note. MSVT = Medical Symptom Validity Test; WAIS-IV = Wechsler Adult Intelligence Scale – Fourth Edition; CVLT-II = California Verbal Learning Test – Second Edition; BVMT-R = Brief Visuospatial Memory Test – Revised; SE = standard error; CAPS-IV = Clinician-Administered PTSD Scale for DSM-IV; PTS = posttraumatic stress; DASS-21 = Depression, Anxiety, and Stress Scale – 21 items; PSQI = Pittsburgh Sleep Quality Index; WHODAS = World Health Organization Disability Assessment Scale.
Models are adjusted for age and education.

Table 6. Adjusted linear regression analyses for 2+ performance validity test failure with symptom validity test failures removed

	Failed 2+ PVTs*					
Covariates	*n*	R^2	*B*	*SE*	*β*	*P*-value
Model 1: Validity-10 Scores≥19 Removed						
Psychiatric						
CAPS-IV PTS Total	449	0.05	26.48	9.04	0.14	.0036
DASS-21 Depression	447	0.02	1.44	0.56	0.12	.0107
DASS-21 Anxiety	447	0.01	0.77	0.50	0.07	.1228
Somatic						
McGill Pain Severity	433	0.02	0.61	0.92	0.03	.5087
PSQI Sleep Disturbance	440	0.03	1.84	1.57	0.06	.2403
WHODAS Functional Impairment	444	0.04	1.32	0.63	0.10	.0361
Model 2: Validity-10 Scores≥23 Removed						
Psychiatric						
CAPS-IV PTS Total	471	0.05	24.09	8.36	0.13	.0041
DASS-21 Depression	469	0.03	1.57	0.52	0.14	.0027
DASS-21 Anxiety	469	0.02	0.96	0.47	0.10	.0407
Somatic						
McGill Pain Severity	452	0.03	1.17	0.84	0.07	.1637
PSQI Sleep Disturbance	462	0.04	2.28	1.43	0.07	.1118
WHODAS Functional Impairment	466	0.05	1.59	0.59	0.12	.0072
Model 3: Validity-10 Scores≥28 Removed						
Psychiatric						
CAPS-IV PTS Total	480	0.07	25.70	7.64	0.15	.0008
DASS-21 Depression	478	0.05	1.65	0.47	0.16	.0005
DASS-21 Anxiety	478	0.04	1.28	0.43	0.14	.0031
Somatic						
McGill Pain Severity	461	0.03	1.22	0.76	0.08	.1118
PSQI Sleep Disturbance	470	0.05	2.79	1.33	0.10	.0363
WHODAS Functional Impairment	475	0.06	1.70	0.54	0.14	.0017

Note. PVT = performance validity test; SE = standard error; CAPS-IV = Clinician-Administered PTSD Scale for DSM-IV; PTS = posttraumatic stress; DASS-21 = Depression, Anxiety, and Stress Scale – 21 items; PSQI = Pittsburgh Sleep Quality Index; WHODAS = World Health Organization Disability Assessment Scale.
*A PVT failure was considered a (1) Medical Symptom Validity Test (MSVT) immediate recognition, delayed recognition, or consistency index ≤85%; or a (2) Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV) reliable digit span score ≤6; or a (3) California Verbal Learning Test – Second Edition (CVLT-II) forced choice score ≤14; or a (4) Brief Visuospatial Memory Test – Revised (BVMT-R) recognition discrimination index score ≤4, recognition hits score ≤4, or percent retained ≤58%.
Models are adjusted for age and education.

disorders (e.g., Major Depressive Disorder, Persistent Depressive Disorder), also known as the deployment trauma phenotype (DTP; Lippa et al., 2015). In the current study, approximately 35% of those who failed 2+ PVTs had DTP, suggesting that these particular comorbid psychiatric conditions may be highly linked to poorer performance validity (Clark et al., 2014; Greiffenstein, & Baker, 2008). Prior research has also suggested that DTP was linked to poorer functional and cognitive outcomes (Amick et al., 2018; Kim et al., 2022; Lippa et al., 2015). Even when adjusted for age and education, 2+ PVTs failures were associated with greater PTS severity and self-reported depression/anxiety (based on self-reported questionnaires) as well as sleep and functional impairment. The only standalone measure, the MSVT, was comparable to the other embedded PVT measures as they were both were associated with negative clinical outcomes (Table 5).

To further ensure that psychiatric factors predicted PVT failure rates, a sensitivity analysis removing those who failed SVTs at all three different cutoffs showed that 2+ PVT failures were associated with greater PTS severity, depression symptoms, and functional impairment. When removing SVT failures at the most conservative cutoff score (≥28; denoting highly probable symptom exaggeration), the PVT failures were additionally linked to increased self-reported anxiety and sleep problems. The results highlight that the relationship between poor PVT performance and psychiatric factors remained in the absence of those who were prone to highly probable symptom exaggeration (Table 6). In sum, findings

https://doi.org/10.1017/S1355617723000711 Published online by Cambridge University Press

indicate that clinicians should consider clinical diagnoses and clinical symptom severity when interpreting validity measures given their strong association with PVT failures.

Past literature showed those who failed the Validity-10 were more likely to also fail PVTs (Jurick et al., 2016). Aase et al. (2021) examined the concordance of the Validity-10 (pass or fail) with embedded PVTs (pass or fail) including the CVLT-II forced choice and total trials 1-5, BVMT-R recognition discrimination score, and CPT-II Commissions score, and found associations at ≥13 and ≥19 cutoff scores (moderate effect sizes), but not at ≥23. In the current study, failure of 2+ PVTs was associated with failure on the Validity-10 on the NSI at ≥19, and ≥23 cutoffs, denoting possible and probable exaggeration, respectively. Approximately 38% of veterans who failed 2+ PVTs also failed the Validity-10 at "possible" exaggeration level. However, 2+ PVT failures were not associated with ≥28 cutoff score, which indicated highly probable exaggeration; this may be attributable to the small sample size ($n = 7$) who met the ≥28 threshold.

The Validity-10 had moderate correspondence in predicting those who failed 2+ PVTs, but low correspondence in predicting those who failed at least one PVT. The latter finding was consistent with Bomyea et al., 2020 which demonstrated that the Validity-10 is a poor predictor of those failing at least one of PVT (e.g., TOMM or CVLT). Several studies have demonstrated that SVTs and PVTs measure separate constructs (Boe & Evald, 2022; Ord et al., 2021) but are related. Therefore, both SVTs and PVTs are essential components in neuropsychological testing and the inclusion of both approaches should be considered.

Failure of SVTs may reflect high clinical distress, a cry for help (Berry et al., 1996; Miskey et al., 2020), and/or psychiatric symptomatology. Specifically, the elevated NSI Validty-10 scores were strongly linked to increased PTS (Aase et al., 2021) and depression symptoms, but not with TBI (Bomyea et al., 2020). Although it may be utilized as a screening tool, the Validity-10 has limitations as it has low sensitivity and not as robust as standalone SVTs (Boone, 2021; Vanderploeg et al., 2014). If failed, clinicians are recommended to follow up utilizing other well-validated SVTs (Lange et al., 2015).

One study showed that having a PTS diagnosis, greater symptom severity, and poorer distress tolerance was associated with failure in the Structured Inventory of Malingered Symptomatology (SIMS), which is a self-reported standalone symptom validity measure (Miskey et al., 2020). They further found that veterans with PTS and depression (which was prevalent in our sample) may have difficulty dealing with strong and negative emotions leading to symptom exaggeration. Furthermore, depression may further contribute to symptom exaggeration as negative cognitive biases may exacerbate symptom report (Agnoli et al., 2023; Armistead-Jehle, 2010; McCormick et al., 2013). Our findings highlight the need for clinical assessments, including the CAPS-4/5 and SCID-4/5, to also include separate validity measures as overreporting can bias findings. Although some studies utilize the SIMS and MMPI which has specific validity indicators (e.g., fake bad scale; Frueh et al., 2000; Miskey et al., 2020), including symptom validity with clinical assessments is not the current standard of care in psychological or psychiatric assessment.

Limitations

The use of the NSI Validity-10 scale as the only SVT is a relative weakness in the study. Future studies should include standalone, well-validated SVT measures as they are robust method in determining response biases. Also, the percent retention from the BVMT-R was used as one of the embedded PVTs included in the analyses; the percent retained does not have a fixed range and can widen based on the amount of information encoded on previous learning trials resulting in highly variable range of scores. Additionally, findings from veteran research sample settings where there are reduced secondary gain of data may not be generalizable to common clinical settings.

Conclusions

Failing of 2+ PVTs may best indicate invalid neuropsychological profiles in a sample of post-9/11 veterans who were informed that their research evaluation would not impact establishing or increasing disability benefits. Failure of PVTs are associated with greater clinical psychiatric diagnoses rather than TBI history. Additionally, PVT failures predicted SVT failure and vice versa. Validity measures are crucial for both neuropsychological testing as well as psychiatric assessments as general practice. Converging data from PVTs and SVTs may be helpful in determining credibility of both neuropsychological evaluations and subjective reports, leading to accurate interpretations and the most appropriate treatments.

Acknowledgments. This work was supported by Translational Research Center for TBI and Stress Disorders (TRACTS), a VA Rehabilitation Research and Development Traumatic Brain Injury National Research Center (B3001-C)

Competing interests. None.

References

Aase, D. M., Soble, J. R., Shepard, P., Akagi, K., Schroth, C., Greenstein, J. E., Proescher, E., & Phan, K. L. (2021). 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

Agnoli, S., Zuberer, A., Nanni-Zepeda, M., McGlinchey, R. E., Milberg, W. P., Esterman, M., DeGutis, J., & Carona, C. (2023). Depressive symptoms are associated with more negative global metacognitive biases in combat veterans, and biases covary with symptom changes over time. *Depression and Anxiety*, *2023*, 1–13. https://doi.org/10.1155/2023/2925551

Amick, M. M., Meterko, M., Fortier, C. B., Fonda, J. R., Milberg, W. P., & McGlinchey, R. E. (2018). The deployment trauma phenotype and employment status in veterans of the wars in Iraq and Afghanistan. *Journal of Head Trauma Rehabilitation*, *33*(2), E30–E40. https://doi.org/10.1097/HTR.0000000000000308

Armistead-Jehle, P. (2010). Symptom validity test performance in U.S. veterans referred for evaluation of mild TBI. *Applied Neuropsychology*, *17*(1), 52–59. https://doi.org/10.1080/09084280903526182

Armistead-Jehle, P., & Hansen, C. L. (2011). Comparison of the repeatable battery for the assessment of neuropsychological status effort index and stand-alone symptom validity tests in a military sample. *Archives of Clinical Neuropsychology*, *26*(7), 592–601. https://doi.org/10.1093/arclin/acr049

Ashendorf, L. (2019). Neurobehavioral symptom validity in U.S. department of veterans affairs (VA) mild traumatic brain injury evaluations. *Journal of Clinical and Experimental Neuropsychology*, *41*(4), 432–441. https://doi.org/10.1080/13803395.2019.1567693

Bailey, K. C., Soble, J. R., Bain, K. M., & Fullen, C. (2018). Embedded performance validity tests in the Hopkins Verbal Learning Test—Revised and the Brief Visuospatial Memory Test—Revised: A replication study. *Archives of Clinical Neuropsychology*, *33*(7), 895–900. https://doi.org/10.1093/arclin/acx111

Benedict, R. (1997). *Brief visuospatial memory test – revised: Professional manual*. Psychological Assessment Resources Inc.

Berry, D. T. R., Adams, J. J., Clark, C. D., Thacker, S. R., Burger, T. L., Wetter, M. W., & Baer, R. A. (1996). Detection of a cry for help on the MMPI-2: An analog investigation. *Journal of Personality Assessment*, *67*(1), 26–36. https://doi.org/10.1207/s15327752jpa6701_2

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, *8*(1), 75–90.

Boe, E. W., & Evald, L. (2022). Symptom and performance validity in neuropsychological assessments of outpatients 15-30 years of age. *Brain Injury*, *37*(3), 1–7. https://doi.org/10.1080/02699052.2022.2158222

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

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, *23*(4), 729–741. https://doi.org/10.1080/13854040802427803

Boone, K. B. (Ed.) (2021). *Assessment of feigned cognitive impairment: A neuropsychological perspective* (2nd ed.). Guilford Publications.

Buysse, D. J., Reynolds 3rd, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213.

Cicerone, K. (1995). The neurobehavioral symptom inventory. *The Journal of Head Trauma Rehabilitation*, *10*(3), 1–17.

Clark, A. L., Amick, M. M., Fortier, C., Milberg, W. P., & McGlinchey, R. E. (2014). Poor performance validity predicts clinical characteristics and cognitive test performance of OEF/OIF/OND veterans in a research setting. *The Clinical Neuropsychologist*, *28*(5), 802–825. https://doi.org/10.1080/13854046.2014.904928

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

Critchfield, E., Soble, J. R., Marceaux, J. C., Bain, K. M., Chase Bailey, K., Webber, T. A., Alex Alverson, W., Messerly, J., Andrés González, D., & O'Rourke, J. J. F. (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

Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, *23*(1), 133–152. https://doi.org/10.1080/13854040701819050

Delis, D., Kaplan, E., & Kramer, G. L. (2001). *D-KEFS examiner's and technical manual*. Pearson Education.

Delis, D. C., Kramer, J., Kaplan, E., & Ober, B. (2000). *California Verbal Learning Test* (2nd ed.). The Psychological Corporation.

Denning, J. H. (2012). The efficiency and accuracy of the Test of Memory Malingering trial 1, errors on the first 10 items of the test of memory malingering, and five embedded measures in predicting invalid test performance. *Archives of Clinical Neuropsychology*, *27*(4), 417–432. https://doi.org/10.1093/arclin/acs044

Denning, J. H., & Shura, R. D. (2017). 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

First, M., Spitzer, R., Gibbon, M., & Williams, J. (1997). *User's guide for the structured clinical interview for DSM-IV axis I disorders-SCID*. American Psychiatric Press.

Fischer, J. E., Bachmann, L. M., & Jaeschke, R. (2003). A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Medicine*, *29*(7), 1043–1051. https://doi.org/10.1007/s00134-003-1761-8

Fortier, C. B., Amick, M. M., Grande, L., McGlynn, S., Kenna, A., Morra, L., Clark, A., Milberg, W. P., & McGlinchey, R. E. (2014). The Boston Assessment of Traumatic Brain Injury – Lifetime (BAT-L) semistructured interview: Evidence of research utility and validity. *The Journal of Head Trauma Rehabilitation*, *29*(1), 89–98.

Frueh, B. C., Hamner, M. B., Cahill, S. P., Gold, P. B., & Hamlin, K. L. (2000). Apparent symptom overreporting in combat veterans evaluated for PTSD. *Clinical Psychology Review*, *20*(7), 853–885. https://doi.org/10.1016/s0272-7358(99)00015-x

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

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen III, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045–1060. https://doi.org/10.1080/02699050110088254

Greiffenstein, M. F., & Baker, W. J. (2008). Validity testing in dually diagnosed post-traumatic stress disorder and mild closed head injury. *The Clinical Neuropsychologist*, *22*(3), 565–582.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(Pt 2), 227–239.

Horner, M. D., Denning, J. H., & Cool, D. L. (2022). Self-reported disability-seeking predicts PVT failure in veterans undergoing clinical neuropsychological evaluation. *The Clinical Neuropsychologist*, *37*(2), 387–401. https://doi.org/10.1080/13854046.2022.2056923

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

Jennette, K. J., Williams, C. P., Resch, Z. J., Ovsiew, G. P., Durkin, N. M., O'Rourke, J. J. F., Marceaux, J. C., Critchfield, E. A., & Soble, J. R. (2022). Assessment of differential neurocognitive performance based on the number of performance validity tests failures: A cross-validation study across multiple mixed clinical samples. *The Clinical Neuropsychologist*, *36*(7), 1915–1932. https://doi.org/10.1080/13854046.2021.1900398 1915.

Jurick, S. M., Twamley, E. W., Crocker, L. D., Hays, C. C., Orff, H. J., Golshan, S., & Jak, A. J. (2016). Postconcussive symptom overreporting in Iraq/Afghanistan Veterans with mild traumatic brain injury. *Journal of Rehabilitation Research and Development*, *53*(5), 571–584. https://doi.org/10.1682/JRRD.2015.05.0094

Kim, S., Currao, A., Bernstein, J., Fonda, J. R., & Fortier, C. B. (2022). Contributory etiologies to cognitive performance in multimorbid post-9/11 veterans: The deployment trauma phenotype. *Archives of Clinical Neuropsychology*, *37*(8), 1699–1709. https://doi.org/10.1093/arclin/acac040

Lange, R. T., Brickell, T. A., Lippa, S. M., & French, L. M. (2015). Clinical utility of the Neurobehavioral Symptom Inventory validity scales to screen for symptom exaggeration following traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *37*(8), 853–862. https://doi.org/10.1080/13803395.2015.1064864

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. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(4), 625–631.

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, *32*(3), 391–421. https://doi.org/10.1080/13854046.2017.1406146

Lippa, S. M., Fonda, J. R., Fortier, C. B., Amick, M. A., Kenna, A., Milberg, W. P., & McGlinchey, R. E. (2015). Deployment-related psychiatric and behavioral conditions and their association with functional disability in OEF/OIF/OND veterans. *Journal of Traumatic Stress*, *28*(1), 25–33. https://doi.org/10.1002/jts.21979

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

McCormick, C. L., Yoash-Gantz, R. E., McDonald, S. D., Campbell, T. C., & Tupler, L. A. (2013). Performance on the Green Word Memory Test

https://doi.org/10.1017/S1355617723000711 Published online by Cambridge University Press

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

McCrea M., Pliskin N., Barth, J., Cox, D., Fink, J., French, L., Hammeke, T., Hess, D., Hopewell, A., Orme, D., Powell, M., Ruff, R., Schrock, B., Terryberry-Spohr, L., Vanderploeg, R., & Yoash-Gantz, R. (2008). Official position of the military TBI task force on the role of neuropsychology and rehabilitation psychology in the evaluation, management, and research of military veterans with traumatic brain injury. *The Clinical Neuropsychologist*, *22*(1), 10–26. https://doi.org/10.1080/13854040701760981

McGlinchey, R. E., Milberg, W. P., Fonda, J. R., & Fortier, C. B. (2017). A methodology for assessing deployment trauma and its consequences in OEF/OIF/OND veterans: The TRACTS longitudinal prospective cohort study. *International Journal of Methods in Psychiatric Research*, *26*(3), e1556. https://doi.org/10.1002/mpr.1556

Melzack, R. (1975). The McGill Pain Questionnaire: Major properties and scoring methods. *Pain*, *1*(3), 277–299.

Meyers, J. E., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology*, *26*(1), 8–15. https://doi.org/10.1093/arclin/acq083

Miskey, H. M., Martindale, S. L., Shura, R. D., & Taber, K. H. (2020). Distress tolerance and symptom severity as mediators of symptom validity failure in veterans with PTSD. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *32*(2), 161–167. https://doi.org/10.1176/appi.neuropsych.17110340

Ord, A. S., Shura, R. D., Sansone, A. R., Martindale, S. L., Taber, K. H., & Rowland, J. A. (2021). Performance validity and symptom validity tests: Are they measuring different constructs? *Neuropsychology*, *35*(3), 241–251. https://doi.org/10.1037/neu0000722

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

Russo, A. C. (2012). Symptom validity test performance and consistency of self-reported memory functioning of Operation Enduring Freedom/Operation Iraqi freedom veterans with positive veteran health administration comprehensive traumatic brain injury evaluations. *Archives of Clinical Neuropsychology*, *27*(8), 840–848. https://doi.org/10.1093/arclin/acs090

Sawyer II, R. J., Testa, S. M., & Dux, M. (2017). Embedded performance validity tests within the Hopkins Verbal Learning Test – Revised and the Brief Visuospatial Memory Test – Revised. *The Clinical Neuropsychologist*, *31*(1), 207–218. https://doi.org/10.1080/13854046.2016.1245787

Schroeder, R. W., & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist*, *25*(3), 437–453. https://doi.org/10.1080/13854046.2011.556668

Schwartz, E. S., Erdodi, L., Rodriguez, N., Ghosh, J. J., Curtain, J. R., Flashman, L. A., & Roth, R. M. (2016). CVLT-II forced choice recognition trial as an embedded validity indicator: A systematic review of the evidence. *Journal of the International Neuropsychological Society*, *22*(8), 851–858. https://doi.org/10.1017/S1355617716000746

Shura, R. D., Miskey, H. M., Rowland, J. A., Yoash-Gantz, R. E., & Denning, J. H. (2016). Embedded performance validity measures with postdeployment veterans: Cross-validation and efficiency with multiple measures. *Applied Neuropsychology: Adult*, *23*(2), 94–104. https://doi.org/10.1080/23279095.2015.1014556

Skinner, H. A., & Sheu, W. J. (1982). Reliability of alcohol use indices. The Lifetime Drinking History and the MAST. *Journal of Studies on Alcohol*, *43*(11), 1157–1170.

Stuss, D. T., Ely, P., Hugenholtz, H., Richard, M. T., LaRochelle, S., Poirier, C. A., & Bell, I. (1985). Subtle neuropsychological deficits in patients with good recovery after closed head injury. *Neurosurgery*, *17*(1), 41–47.

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

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285–1293.

Tiffen, J. (1968). *Purdue pegboard examiner's manual*. London House.

Traumatic Brain Injury Center of Excellence (TBICoE). (2021, October 14). *DOD TBI Worldwide Numbers*. https://www.health.mil/Military-Health-Topics/Centers-of-Excellence/Traumatic-Brain-Injury-Center-of-Excellence/DOD-TBI-Worldwide-Numbers

Üstün, T. B., Kostanjsek, N., Chatterji, S., & Rhehm, J. (Eds.) (2010). *Measuring health and disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0)*. WHO Press.

Vanderploeg, R. D., Cooper, D. B., Belanger, H. G., Donnell, A. J., Kennedy, J. E., Hopewell, C. A., & Scott, S. G. (2014). Screening for postdeployment conditions: Development and cross-validation of an embedded validity scale in the Neurobehavioral Symptom Inventory. *Journal of Head Trauma Rehabilitation*, *29*(1), 1–10. https://doi.org/10.1097/HTR.0b013e318281966e

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, *23*(2), 297–313. https://doi.org/10.1080/13854040802232682

Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV Digit Span indices for identifying noncredible performance validity among cognitively impaired and unimpaired examinees. *Clinical Neuropsychologist*, *32*(4), 657–670. https://doi.org/10.1080/13854046.2017.1415374

Wechsler, D. (2001). *Wechsler Test of Adult Reading: WTAR*. The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale (manual)* (4th ed.). Psychological Corporation.

Young, J. C., Roper, B. L., & Arentsen, T. J. (2016). Validity testing and neuropsychology practice in the VA healthcare system: Results from recent practitioner survey. *The Clinical Neuropsychologist*, *30*(4), 497–514. https://doi.org/10.1080/13854046.2016.1159730