



Published in final edited form as:

Neuropsychology. 2023 January ; 37(1): 93–103. doi:10.1037/neu0000857.

Varying Failure Criteria on Performance Validity Tests Influences Interpretation of Cognitive Outcomes

Savanna M. Tierney^{1,*}, Zinat Taiwo^{1,3}, Tabina K. Choudhury¹, Sara M. Lipka², Maya Troyanskaya^{1,3}, Nicholas J. Pastorek^{1,3}, Brian Miller^{1,3}, Jennifer Romesser⁴, Anita Sim⁵

¹Michael E DeBakey VA Medical Center, Houston, TX, USA

²Walter Reed National Military Medical Center, Bethesda, MD, USA

³H. Ben Taub Department of Physical Medicine and Rehabilitation, Baylor College of Medicine, Houston, TX, USA

⁴VA Salt Lake City Health Care System, Salt Lake City, UT, USA

⁵Minneapolis VA Health Care System, Minneapolis, MN, USA

Abstract

Objective: The current study examined the effects of applying various performance validity test (PVT) failure criteria on the relationship between cognitive outcomes and posttraumatic stress (PTS) symptomology.

Method: One hundred and ninety-nine veterans with a history of mild traumatic brain injury referred for clinical evaluation completed cognitive tests, PVTs, and self-report measures of post-traumatic stress symptoms and symptom exaggeration. Normative T-scores of select cognitive tests were averaged into memory, attention/processing speed, and executive functioning composites. Separate one way analyses of variance assessed differences among high PTS (n = 140) versus low PTS (n = 59) groups and were repeated excluding participants based on varying combinations of PVT failure criteria.

Results: When no PVTs were considered, the high PTS group demonstrated worse performance across all three cognitive domains. Excluding those who failed two or more standalone, or two or more embedded validity measures resulted in group differences across all cognitive composites. When participants were excluded based on failure of any one embedded and any one standalone PVT measure combined, the high PTS group performed worse on the executive functioning and attention/processing speed composites. The remaining three proposed methods to control for performance validity resulted in null PTS - cognition relationships. Results remained largely consistent after controlling for symptom exaggeration.

Conclusions: Methods of defining PVT failure can greatly influence differences in cognitive function between groups defined by PTS symptom levels. Findings highlight the importance of considering performance validity when interpreting cognitive data and warrant future investigation of PVT failure criteria in other conditions.

*Corresponding Author: Tierneysavanna@gmail.com.

Keywords

performance validity tests; cognition; post-traumatic stress

Clinical neuropsychology continues to evolve towards optimally valid and reliable assessment of cognitive and emotional functioning, to assist in etiological clarification, treatment planning, and functional prognosis of cognitive disorders. However, the integrity and utility of a neuropsychological evaluation can be considerably limited when results of cognitive testing are determined to be invalid. Variable and/or inadequate effort exerted during the evaluation can reduce the reliability of cognitive data and is usually assessed via performance validity tests (PVTs) which indicate whether results depict an accurate representation of an examinee's true abilities (Larrabee, 2012). Negative impression management (i.e., over-reporting or magnification of cognitive, emotional, and physical symptoms) has also been identified as a form of response bias and is evaluated using symptom validity tests (SVTs; Larrabee, 2012).

The need for PVTs as part of a comprehensive neuropsychological evaluation has been well articulated (Bush et al., 2005; Heilbronner et al., 2009); however, establishing standardized protocols for implementing such procedures (including but not limited to the types, number, and order of measures used) has proven to be a challenge. A 2015 survey of neuropsychologists found that nine out of ten neuropsychologists utilize embedded or standalone PVTs and administer four to six PVTs, on average (Martin et al., 2015). However, in the event of discrepant PVT performance, there was variability in criteria for identifying suboptimal engagement and procedural steps. For example, respondents varied in their threshold for determining performance invalidity: 35% used a failure of two PVTs, 27% combined one failed PVT measure and one indicator of poor engagement outside of validity testing (e.g. behavioral observations); and 10% of the sample reported reliance on clinical judgement, failure of one well-established PVT, or a ratio of failed PVTs to number administered as the criterion for identifying performance invalidity. Further, interpretation of cognitive results varied depending on the type of PVT administered (i.e., embedded versus standalone), such that when results from standalone and embedded validity measures were incongruous, neuropsychological test data were more likely to be considered unreliable if standalone PVTs were failed but embedded measures were passed, compared to when the inverse was true (54% versus 23%; Martin et al., 2015).

When multiple PVTs are administered, it is most common for neuropsychologists to conclude that cognitive data is unreliable when there is failure of 2 PVTs (Martin, Schroeder, & Odland, 2015; Schroeder & Marshall, 2011). Although many studies support the finding that 2 PVT failures results in excellent specificity (e.g. Chafetz, 2011; Larrabee, 2003; Meyers & Volbrecht, 2003; Sollman, Ranseen, & Berry, 2010), failure of one PVT may convey important information in certain circumstances (Proto et. al., 2014). Lippa (2018) concluded that when PVTs have been validated in samples similar to the population of interest (and in the absence of significant neuropathology or other explanation for PVT failures), clinicians should feel confident in their determination of invalid responding in many cases when a single PVT is failed and in most cases when two or more PVTs

are failed. Gaining a clearer understanding of the effects of specific PVT failure criteria may have considerable implications for research on the relationship between cognition and various neurologic and psychiatric conditions, especially where there is reason to suspect higher base rates of performance invalidity.

The relationship between Post-Traumatic Stress Disorder (PTSD) and cognition has received a great deal of attention and provides an important opportunity to explore the implications of using various PVT failure criterion. Many studies have found that individuals with PTSD perform significantly worse on measures of memory than those without (for a review, see Vasterling & Brailey, 2005). These findings appear to persist after controlling for the influence of attentional difficulties (Johnsen, Kanagaratnam, & Asbjørnsen, 2008; Yehuda, Golier, Tischler, Stavitsky, & Harvey, 2005). Other studies have concluded that PTSD is also associated with poorer performance on tasks of immediate and sustained attention (Vasterling et al., 2002), as well as measures of working memory (Brandes et al., 2002; Jenkins et al, 1998, Vasterling et al., 1998, 2002), and phonemic and semantic fluency (Gil, Calev, Greenberg, Kugelmass, & Lerer, 1990). The topic of test validity in these studies, however, was largely unaddressed. It has been suggested that failure to formally assess respondent validity may, in some cases, lead to the erroneous impression that PTSD is associated with cognitive impairment and neurological changes (Rosen, 2006).

Addressing this concern, Sullivan and colleagues (2003) applied a cut-score of 45 on Trial 1 from the Test of Memory Malingered (TOMM) and found no significant differences in cognitive performance among veterans with and without PTSD diagnoses. Similarly, Demakis and colleagues (2008) reported that the severity of PTSD symptoms among their large sample of Canadian litigants did not have any effect on cognitive ability after individuals who failed PVTs were removed. In contrast, Marx et al., 2009 reported a relationship between cognition and PTSD severity among a sample of active duty Army members; however, it should be noted the authors applied a cut-score of 38 for the TOMM Trail 1 to determine invalid performance, which may have resulted in an increased number of false negatives among this relatively young (mean reported age of 25 years), independently functioning sample (Denning, 2012; Martin et al., 2020).

The comorbidity between PTSD and Traumatic Brain Injury (TBI) has also received increased attention. A 2010 examination of 13,201 US military administrative records found that Veterans with positive TBI screens and confirmed TBI status were more likely than those without confirmed TBI status to have diagnoses of PTSD, anxiety, and adjustment disorders (Carlson, et al., 2010). Performance invalidity appears to at least partially account for cognitive findings among these clinical groups. An investigation of 142 active-duty service members evaluated following suspected or confirmed history of mild traumatic brain injury (46% with comorbid anxiety/PTSD diagnoses) found that PVT results accounted for the most variance in cognitive test scores, above demographic, concussion history, symptom validity, and psychological distress variables (Armistead-Jehle, Cooper, & Vanderploeg, 2015). Furthermore, Wisdom and colleagues (2013) examined possible cognitive impairments correlated with PTSD among veterans with a history of mild traumatic brain injury after excluding individuals who failed a single, well-validated PVT. Results revealed that failure on the Word Memory Test (WMT) was associated with

significantly poorer performance on almost all cognitive tests administered. Additionally, no significant differences were detected between individuals with and without PTSD symptoms after controlling for suboptimal effort (i.e., failure on the WMT).

Despite recognizing PVT failure is associated with significant declines in cognitive testing performance, the question remains as to how research and clinical data should be interpreted in the context of these findings, especially among samples where neurologic injury (and therefore, cognitive dysfunction directly related to neurologic injury) is believed to be relatively limited. While ample research has examined the classification accuracy of PVTs and their cutoff criteria (e.g., at least two failures) in various samples, we are not aware of any studies that have examined the influence of varying PVT failure criteria on the interpretation of clinical relationships often assessed in neuropsychological evaluation (e.g., post-traumatic stress symptoms and cognition).

The purpose of the current study was to examine the effects of differing but commonly used PVT failure criteria on the relationship between Post-Traumatic Stress (PTS) symptomology and cognitive domains thought to be impacted in PTSD - specifically, attention/processing speed, memory, and executive functioning. The relationship between cognition and PTS symptoms was examined for demonstrative purposes among a sample of veterans with history of mild traumatic brain injury (TBI), though the results should have broader implications for other health conditions, especially those associated with elevated rates of symptom invalidity. It was hypothesized that when no PVTs were considered, significant relations would be detected between cognition and PTS symptom level. Broadly speaking, as PVT requirements varied by administration method (e.g., embedded versus standalone) and number required to meet failure criteria (e.g., one versus two or more), it was hypothesized that the observed relationship between PTS symptom level and cognition would vary as well. We speculated that validity groupings with the most “strict” failure criteria (e.g., PVT classification schemes that exclude relatively more participants) would result in the greatest exclusion of participants and would be less likely to demonstrate significant relations between PTS symptom level and cognition. In contrast, validity groupings with “lenient” failure criteria (e.g., PVT classification schemes that exclude relatively less participants) would result in the least exclusion of participants and therefore would be more likely to demonstrate cognitive differences between individuals with low and high PTS symptom levels (similar to incorporating no PVT).

Methods

Participants

The original study sample included 261 (Bodapati et al., 2019) veterans with a history of mild TBI (mTBI) referred for clinical treatment and neuropsychological evaluation at one of four Veterans Affairs (VA) hospitals and medical centers located in the Western, Southern, and Midwestern United States. All participants completed a standard neuropsychological assessment battery, which included administration of standalone (TOMM, WMT, Rey Memory for Fifteen Items Test (MFIT)) and embedded PVTs (California Verbal Learning Test-Second Edition (CVLT-II: Forced Choice (FC), Wisconsin Card Sorting Test (WCST): Failure to Maintain Set (FTMS), Wechsler Adult Intelligence Scale -IV (WAIS-IV):

Reliable Digit Span (RDS)), PTSD Checklist-Civilian version (PCL-C) and the Minnesota Multiphasic Personality Inventory-2 (MMPI-2). This study was approved by the Institutional Review Boards of all participating entities and informed consent was obtained prior to the evaluation. All assessments (or evaluations) were performed by licensed psychologists or trainees under the direct supervision of licensed psychologists and occurred between May 2011 and December 2013.

Information regarding loss of consciousness (LOC), confusion, or posttraumatic amnesia (PTA) was obtained through a semi-structured interview modeled after the VA Comprehensive TBI Evaluation (Belanger, Uomoto, & Vanderploeg, 2009), which was then translated into TBI severity based on the American Congress of Rehabilitation Medicine (American College of Rehabilitation Medicine, 1993) guidelines (Bodapati et al., 2019; Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation, 1993). Veterans with a history of mTBI were included in the study, as indicated by an external force to the head followed by 30 min or less of LOC, 24 hours or less of PTA, or 24 hours or less of alteration in mental state. On average, the time between most recent mTBI and clinical assessment was 63 months within the total sample. There were no significant differences in the time from last reported injury between the high PTS and low PTS symptom groups. Participants whose self-report did not meet criteria for mTBI ($n=41$) as well as participants with incomplete MMPI-2 profiles ($n=4$), or evidence of inconsistent or fixed responding (failure of Variable Response Inconsistency; VRIN and True Response Inconsistency; TRIN) scales ($n=7$) were excluded from this investigation. Data from participants missing any of the required PVT measures were also excluded ($n=10$). The final study sample included 199 participants. The PCL-C (DSM-IV version; Weathers, Litz, Huska, & Keane, 1994) was used to characterize high PTS and low PTS symptom groups using a cut-score of 50 (e.g., Kang et. al., 2003, Tanielian & Jaycox, 2008). Veterans who endorsed symptoms at a score of 50 or higher on the PCL-C were included in the high PTS group ($n=140$) and those scoring 49 or lower were classified as the low PTS group ($n=59$). There were no significant groups differences across sex, ethnicity, branch of service, education, single word-reading scores or marital, student, or occupational status. The low PTS ($6/5/2022M=33.47$, $SD=8.05$) group was slightly older than the high PTS group ($M=31.30$, $SD=6.39$; $p=0.04$). Importantly, age adjusted normed scores were used in composites and data analysis to minimize concerns for age effects. Participant demographic data is shown in Table 1.

Measures

Cognitive functioning: All veterans were administered a standardized neuropsychological battery. The Wide Range Achievement Test 4 (WRAT-4; Wilkinson & Robertson, 2006) was included as a measure of global premorbid cognitive functioning (see Table 1). Normative T-scores of select neuropsychological tests were averaged into three composites of cognitive domains commonly thought to be affected by PTSD. Measured domains and associated tests included: (1) memory - assessed via Trials 1–5 raw sum (Total Learning, or TL), short delay free recall (SD), and long-delay free recall (LD), from the CVLT-II (Delis, Kramer, Kaplan, & Ober, 2000) (2) attention/processing speed – assessed via the WAIS-IV coding, symbol search, and digit span subtests (Wechsler, 2008) and Trail

Making Test (TMT) Part A (Reitan & Wolfson, 1985) and (3) executive functioning – assessed via TMT Part B, Stroop Color-Word Interference index (Golden & Freshwater 1978; Stroop, 1935), and total score for letter fluency (FAS; Gladsjo et al., 1999). Of note, cognitive data for the CVLT-II was normed using the CVLT-II manual (Delis, Kramer, Kaplan, & Ober, 2000; age and sex). Cognitive data for the WAIS-IV subtests was normed using the WAIS-IV manual (Wechsler, 2008; age). Trail Making Test and Total Letter Fluency were normed using Heaton Comprehensive Norms (Heaton et al., 1991; age, sex, race, education). Stroop data was normed with the manual (Golden & Freshwater 1978; education and age).

Performance validity: The present study included a total of six standalone and embedded performance validity measures. Standalone PVTs included the WMT (Green, 2003; Green, Allen, & Astner, 1996), TOMM (Tombaugh, 1996), and the MFIT (Rey, 1964) recognition trial. Embedded PVTs within the current study included the WCST (Heaton, Chelune, Talley, Kay, & Curtiss, 1993) FTMS, WAIS-IV RDS (Greiffenstein, Baker, & Gola, 1994), and the CVLT-II FC. Recommended manual-based cut scores from primary scales/trials were used for the WMT. Cutoff values for the TOMM (Trial 1 = 41 or less; Martin et al., 2020), MFIT free recall and recognition combination score (Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002), RDS (6 or less; Greve, Bianchini, & Brewer, 2013; Meyers & Volbrecht, 2003), WCST (FTMS = 3 or more; Greve & Bianchini, 2007; Greve, Heinly, Bianchini, & Love, 2009), and CVLT-II FC raw score (14 or less; Root, Robbins, Chang, & Van Gord, 2006) were based on published values aimed at maximizing specificity. Varying combinations of these measures were used to form “failure criteria” groups modeled after ways in which researchers and/or clinicians may choose to exclude “PVT fails” in practice.

Six groups defined by various PVT failure patterns were considered: (1) any standalone – classified as failure on at least one of the identified standalone PVTs (either WMT, TOMM Trial 1, or MFIT), (2) two or more standalone – classified as failure on two or more of the identified standalone PVTs (either WMT, TOMM Trial 1, or MFIT), (3) any embedded – classified by c (either RDS, WCST FTMS, or the CVLT FC), (4) two or more embedded – classified as failure on two or more of the identified embedded PVTs (either RDS, WCST FTMS, or the CVLT FC), (5) combination of at least one embedded and one standalone – classified as failure on at least one of the embedded and one of the standalone PVTs, and (6) any combination of two or more – classified by failure on two or more of the identified PVTs (embedded or standalone).

Data Analysis

Separate one way analyses of variance (ANOVA) were used to assess differences among the high PTS ($n = 140$) versus low PTS symptom groups ($n = 59$) across three composites of cognitive domains thought to be impaired among individuals with PTSD: memory, attention/processing speed, and executive functioning. Analyses were repeated, excluding participants based on combinations of “failure criteria” described above. Additionally, effect sizes (Cohen’s d) of the various significant and non-significant between-group differences on cognitive testing performances for each of the six PVT failure levels were calculated based on group means and SDs .

As over-reporting may influence characterization of the PTS symptom groups determined by the PLC-C, analyses were rerun with consideration of scores on the MMPI-2-RF Infrequent Responses (F-r) scale. The F-r scale serves as a general over-reporting measure and comprises 32 items that are rarely endorsed in the MMPI-2-RF normative sample (i.e., were answered in the keyed direction by 10% or less). As the scale has demonstrated utility in assessing symptom misrepresentation (Sellbom and Bagby, 2010, Sellbom et al., 2010), participants with a T-score of 120 on the F-r scale (which was felt to convey a high likelihood of symptom invalidity) were excluded based on standard interpretive guidelines. This set of analyses included 117 adults in the high PTS group and 46 adults in the low PTS group.

Transparency and Openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow Journal Article Reporting Standards (Kazak, 2018). All data and materials (as well as software application or custom code) support the claims herein and comply with field standards. Data sharing may be requested (but cannot be guaranteed, based on institutional policies): The data that support the findings of this study may be available upon submission of a reasonable request and the receipt of necessary approvals. Data were analyzed using JMP, version 14.0.0 statistical program. This study's design and its analysis were not pre-registered.

Results

When no PVTs were considered, results of one-way ANOVA suggest that individuals with high PTS symptomology performed significantly worse across all three cognitive domains (all p s < .05). Similarly, excluding those who have failed two or more standalone validity measures, as well as two or more embedded validity measures, resulted in significant group differences such that individuals in the high PTS group performed worse across all three cognitive composites compared to participants in the low PTS symptom group (all p s < .05). Mean cognitive composite scores for memory, executive functioning, and attention/processing speed across PVT groups are shown in Figures 1, 2, and 3, respectively.

Excluding individuals based on the combination of failures on at least one embedded and at least one standalone PVT measure resulted in group differences on tasks of executive functioning ($p = .048$) and attention/processing speed ($p = .043$), such that individuals in the high PTS group performed significantly worse compared to veterans in the low PTS group. Scores on the memory composite did not significantly differ when this PVT failure criteria was applied ($p > .05$). There were no group differences observed in the three cognitive domains when participants were excluded on the basis of any standalone PVT failure, any combination of two or more failed PVTs (embedded or standalone), or any embedded PVT failures (all p s > .05). Failure rates across performance validity measure and group are displayed in Table 2.

Cohen's d effect sizes were calculated to assess possible influence of sample size. When no PVTs were considered, findings revealed medium effects of high PTS group membership on memory, executive functioning, and attention/processing speed composites. When

performance validity failure was considered, effects of group varied from small to medium across the three cognitive domains (Cohen's d range: 0.08–0.49). Of note, when individuals failing two or more embedded, or two or more standalone PVTs were excluded (resulting in varying total number of participants included in analyses), medium sized effects of group on executive functioning and attention/processing speed measures indicate that findings are not merely a factor of sample size alone. Table 3 displays Cohen's d effect sizes by group across cognitive domain and exclusion criteria.

After controlling for possible symptom overreporting on the PCL-, analyses performed on the subset of participants ($N = 163$) with valid MMPI-2-RF F-r scores revealed largely similar results. Relative to the sample not controlling for symptom overreporting, there were no group differences in executive functioning scores when participants were excluded based on the failures of at least one embedded and at least one standalone PVT combined. There were no differences between groups observed on memory performance when participants were excluded on the basis of 2 or more embedded, or 2 or more standalone PVT failures. All other findings remained consistent.

Discussion

The current study examined the effects of different, commonly used PVT failure criteria on the relationship between cognitive outcomes and health conditions (i.e., PTS symptoms) among a sample of veterans with history of mTBI. Our initial analyses examined the relationship of memory, attention/processing speed, and executive functioning without taking into account performance validity. When no PVTs were considered, our results were consistent with several previous studies suggesting that PTSD is associated with observable cognitive weaknesses (for a review, see Vasterling & Brailey, 2005). Scott and colleagues (2015) presented the first systematic meta-analysis of neurocognitive outcomes associated with PTSD. The report was based on data from 60 studies totaling 4,108 participants, including 1,779 with PTSD, 1,446 trauma-exposed comparison participants, and 895 healthy comparison participants without trauma exposure. In line with current study results when not considering PVT failure, analyses revealed significant neurocognitive effects associated with PTSD, across measures of verbal learning, speed of information processing, attention/working memory, and verbal memory. It should be noted however, these investigations often highlight the lack of PVT inclusion as a limitation (e.g., Scott et al., 2015). Indeed, the totality of findings in the current study support this concern.

When controlling for performance validity on the basis of any standalone PVT failure, any combination of two or more failed PVTs, or any embedded PVT failures, there were no differences in cognitive performance between the high PTS and low PTS symptom groups. Further, Cohen's d effect size calculations revealed consistently small-sized effects associated with these analyses suggesting that null results are not attributable to sample size alone. For example, effect sizes when participants were excluded on the basis of any combination of 2 or more PVT failures were slightly larger than those when excluding for any embedded PVT failure, despite the latter grouping retaining more participants. Of course, the argument could be made that true cognitive deficits in the high PTS group increased the likelihood of false positive findings on individual PVTs, although it

is important to note that the participants in the current sample were relatively young and living in the community, thus mitigating concerns that genuine cognitive impairment was responsible for elevated false positive findings on PVTs in the high PTS group. In total, 3 of the 6 proposed methods to control for performance validity resulted in null PTS-cognition relationships. Consistent with Demakis, Gervais, and Rohling, (2008), these particular findings suggest that observed differences in cognitive performance among individuals with low and high PTS symptoms may be potentially mitigated by controlling for performance validity failure.

Just as in practice, however, our results present challenging nuances to consider. Although requiring failure on two or more embedded or two or more standalone validity measures may be perceived as strict, and thus an effective control against performance invalidity, these groupings allowed for participants who failed one PVT to remain in the analyses. Generally speaking, when 2 PVT failures were required for exclusion within the current sample, we see robust differences between our high PTS and low PTS groups across cognitive composites (i.e., two or more embedded, two or more standalone, the combination of any embedded and any standalone PVT failure). The exception to this was the grouping in which participants were excluded based on the failure on any combination of 2 or more PVTs (standalone or embedded) wherein consistently null differences in cognitive performance were observed. Importantly, this category reflected a failure of any 2 of the 6 total PVTs included in this study. While defining possible suboptimal test engagement as failing two or more PVTs is common practice, some have cautioned that failure on even one PVT should elicit consideration for performance invalidity, particularly in individuals with histories of mTBI (e.g., Lippa, 2018, Proto et al., 2014). The pattern of results in the current study appears consistent with this line of thought. When participants were excluded based on the failure of a single PVT (i.e., any standalone or any embedded measure), results show no significant differences in cognitive performances between groups.

Perhaps the most clinically important finding in the current study is the discrepant outcomes observed when applying criteria that requires failure on two or more standalone PVT measures and exclusion based on the combination of failure on any 2 or more PVTs (standalone or embedded). Application of one criterion results in significant differences in cognitive performance across all three composites and the other yields null cognitive differences between groups. We found this striking, as both methods of controlling for performance validity appear reasonable and are consistent with the prevailing wisdom of requiring failure of two or more PVTs to identify suboptimal engagement. Certainly, if an investigation reported either of these criteria within their methodology most researchers and clinicians would be satisfied with the approach. Yet, the conclusions drawn from these results differ greatly. Interestingly, discrepant outcomes were also observed when applying criteria that requires failure on two or more standalone PVT measures versus any embedded PVT failure. In this case, both criteria result in a similar number of participants excluded, from similar group distributions (i.e., rates of exclusion among high PTS and low PTS groups). Yet again, one criteria results in significant differences in cognitive performance across all three composites and the other yields null cognitive differences between groups. These findings highlight the drastic impact that decisions about how performance validity failure is defined may have on interpretation of cognitive data.

The current study is not without limitations. The predominantly male, veteran based sample, who presented for evaluation at a TBI clinic, limits generalizability of our results. Additionally, participants in the current sample were individuals seeking specialty TBI services, which likely accounts for the high failure rates of PVTs and SVTs in the current sample (e.g., Tombaugh, 1996; Young, Roper, & Sawyer, 2011; Young, Sawyer, Roper, & Baughman, 2012). While high rates of PVT failure were ideal for examining the impact of varying PVT failure criteria, researchers and clinicians are encouraged to use caution when generalizing findings from this unique sample. All veterans in the present study had a history of at least one concussion. However, this was not a threat to internal validity as injuries in the totality of the sample were characterized as mild in nature so there were no differences in injury characteristics between groups. We also recognize that the choice of PVT may have influenced outcomes; it is possible that tasks such as the TOMM and WMT, which at face value, appear memory-focused may have resulted in higher rates of exclusion for those with memory complaints and thus, lessened the likelihood of identifying memory differences on testing. Similarly, inclusion of additional PVTs in the domains of attention/processing speed and executive functioning may have reduced effect sizes in these domains. It is worth noting that although there were statistically significant group differences when some PVT failure criteria were applied, these represented relatively modest normative score differences and the normative scores associated with the high PTS group were consistently in the average range.

Additionally, self-reported PTSD symptomology on the PCL-C was used to divide the groups, rather than a clinician-administered structured interview. Although the PCL-C elevations reported by the high PTS group were clinically meaningful (and nearly three standard deviations higher than the low PTS group), it is possible that some of the veterans would not have met formal criteria for a PTSD diagnosis. It should also be noted that participants were not instructed to think about the mTBI incident during the administration of the PCL-C. As such, we are not able to speak to the chronicity of participant's PTS symptoms, which should be considered when attempting to generalize the current study findings. In addition, the MMPI-2-RF F-r scale was used as a proxy for potential symptom exaggeration on the PCL-C, although we are unaware of any studies that have empirically validated this assumption. Lastly there are several unmeasured health-related conditions that could possibly affect cognition among veteran populations (e.g., exposure to burn pits) that were not evaluated as part of the current investigation.

Of note, there are a variety of reasons why differences in the manner in which performance invalidity has been addressed does not entirely explain the correlation between PTSD symptoms and cognitive deficits reported in extant literature. Vasterling and Brailey (2005) highlight consistent material specific cognitive changes in PTSD (e.g., verbal memory declines as opposed to visual memory; specific aspects of reduced attention), parallel findings from neurobiological models, and observed weaknesses in general clinical settings without identifiable secondary gain in support of the complex impacts PTSD may have on cognition. Further, they emphasize that "PTSD does not occur in a vacuum and is associated with psychiatric comorbidities, adverse health consequences, and somatic insult at the time of traumatization, each of which could potentially exert adverse consequences on neuropsychological functioning" (Vasterling and Brailey, 2005, p. 192). Still, the dearth

of studies examining PTS (or other clinical syndromes for that matter) and cognition that incorporate PVTs in their conceptualization (to a greater extent than noting the absence as a limitation) is problematic given the largely similar performance on cognitive measures observed between veterans with high and low PTS symptoms depending on how performance invalidity was defined in the current study.

Neuropsychologists are ethically and professionally bound to provide accurate diagnostic impressions and recommendations. Studies suggest the ability of clinicians and researchers to determine the validity of findings based only on subjective clinical judgment without the use of performance and symptom validity tests is weak, inaccurate, and subject to bias (Bush & Morgan, 2012; Guilmette, 2013). Failure to assess for response bias may result in a variety of consequences including (but not limited to): erroneous diagnoses and possible iatrogenic effects, further exhaustion of already limited resources (e.g., medical appointments, monetary compensation) and increased distrust of specific clinical populations (Howe, 2012). In settings where cognitive performance dictates treatment (e.g., epilepsy monitoring units), mischaracterization of impairments may have significant health consequences. Similarly, misattributing poor performance to a diagnosed condition (e.g., PTSD, mTBI) rather than identifying performance invalidity may limit consideration of critical behavioral recommendations (e.g., sleep hygiene, pain management, cognitive strategies). To our knowledge, this is the first study to strategically examine how varying PVT failure criteria influences the interpretation of cognitive data. Rather than clear-cut recommendations for PVT failure criteria, the findings highlight how selection of particular failure criteria may drastically change interpretation of important clinical outcomes. Future studies should continue to evaluate PVT failure criteria across other settings and patient populations.

Acknowledgements

This material is based upon work supported in part by: the United States (U.S.) Department of Veterans Affairs Merit Review Awards Numbers I01 CX001820 (PI Scheibel, RS). The views in this article are those of the authors and do not reflect the official policy or position of the Department of Defense, the Veterans Affairs Administration, nor the United States Government. The authors thank the study volunteers for their participation.

All data and materials (as well as software application or custom code) support the claims herein and comply with field standards. Data sharing may be requested (but cannot be guaranteed, based on institutional policies): The data that support the findings of this study may be available upon submission of a reasonable request and the receipt of necessary approvals. Data were analyzed using JMP, version 14.0.0 statistical program. This study's design and its analysis were not pre-registered.

References

- American Congress of Rehabilitation Medicine. (1993). Definition of mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 8, 86–87.
- Armistead-Jehle P, Cooper DB, & Vanderploeg RD (2016). The role of performance validity tests in the assessment of cognitive functioning after military concussion: A replication and extension. *Applied Neuropsychology: Adult*, 23(4), 264–273. [PubMed: 26569384]
- Belanger HG, Uomoto JM, & Vanderploeg RD (2009). The Veterans Health Administration system of care for mild traumatic brain injury: Costs, benefits, and controversies. *The Journal of Head Trauma Rehabilitation*, 24, 4–13. 10.1097/HTR.0b013e3181957032 [PubMed: 19158591]
- Bodapati AS, Combs HL, Pastorek NJ, Miller B, Troyanskaya M, Romesser J, Sim A, & Linck J (2019). Detection of symptom over-reporting on the Neurobehavioral Symptom Inventory in

- OEF/OIF/OND veterans with history of mild TBI. *The Clinical Neuropsychologist*, 33, 539–556. 10.1080/13854046.2018.1482003 [PubMed: 29865932]
- Boone KB, Salazar X, Lu P, Warner-Chacon K, & Razani J (2002). The Rey 15-item recognition trial: a technique to enhance sensitivity of the Rey 15-item memorization test. *Journal of Clinical and Experimental Neuropsychology*, 24, 561–573. 10.1076/jcen.24.5.561.1004 [PubMed: 12187441]
- Brandes D, Ben-Schachar G, Gilboa A, Bonne O, Freedman S, & Shalev AY (2002). PTSD symptoms and cognitive performance in recent trauma survivors. *Psychiatry Research*, 110, 231–238. 10.1016/s0165-1781(02)00125-7 [PubMed: 12127473]
- Bush SS, Ruff RM, Tröster AI, Barth JT, Koffler SP, Pliskin NH, Reynolds CR, & Silver CH (2005). Symptom validity assessment: practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, 20, 419–426. 10.1016/j.acn.2005.02.002 [PubMed: 15896556]
- Bush SS, & Morgan JE (2012). Improbable neuropsychological presentations: Assessment, diagnosis, and management. In Bush SS (Eds.), *Neuropsychological practice with veterans* (pp. 27–43). Springer Publishing Company.
- Chafetz M (2011). Reducing the probability of false positives in malingering detection of Social Security disability claimants. *The Clinical Neuropsychologist*, 25, 1239–1252. 10.1080/13854046.2011.586785 [PubMed: 21722051]
- Delis DC, Kramer JH, Kaplan E, & Ober BA (2000). *California Verbal Learning Test*, 2nd Ed. San Antonio, TX: Psychological Corporation.
- Demakis GJ, Gervais RO, & Rohling ML (2008). The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *The Clinical Neuropsychologist*, 22, 879–895. 10.1080/13854040701564482 [PubMed: 18756390]
- Denning JH (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, 417–432. 10.1093/arclin/acs044 [PubMed: 22543569]
- Gil T, Calev A, Greenberg D, Kugelmass S, & Lerer B (1990). Cognitive functioning in Post-Traumatic Stress Disorder. *Journal of Traumatic Stress*, 3, 29–45. 10.1002/jts.2490030104
- Gladsjo JA, Schuman CC, Evans JD, Peavy GM, Miller SW, & Heaton RK (1999). Norms for letter and category fluency: demographic corrections for age, education, and ethnicity. *Assessment*, 6, 147–178. 10.1177/107319119900600204 [PubMed: 10335019]
- Golden CJ, & Freshwater SM (1978). Stroop color and word test.
- Green P, Allen L, Astner K (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms*. Durham, NC: CogniSyst.
- Green P (2003). *Green's word memory test for windows: User's manual*. Edmonton, AB: Green's Publishing.
- Greiffenstein MF, Baker WJ, Gola T (1994) Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6, 218–224.
- Greve KW, Bianchini KJ (2007). Detection of cognitive malingering with tests of executive functioning. In: Larrabee GJ (Eds.), *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.
- Greve KW, Heinly MT, Bianchini KJ, & Love JM (2009). Malingering detection with the Wisconsin Card Sorting Test in mild traumatic brain injury. *The Clinical Neuropsychologist*, 23, 343–362. 10.1080/13854040802054169 [PubMed: 18609328]
- Greve KW, Bianchini KJ, & Brewer ST (2013). The assessment of performance and self-report validity in persons claiming pain-related disability. *The Clinical Neuropsychologist*, 27, 108–137. 10.1080/13854046.2012.739646 [PubMed: 23157188]
- Guilmette T (2013). The role of clinical judgment in symptom validity assessment. In Carone D & Bush S (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 31–44). New York, NY: Springer
- Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation: Definition of mild traumatic brain injury. *J Head Trauma Rehabil* 1993; 8: pp. 86–87.

- Heaton RK, Chelune GJ, Talley JL, Kay GG, Curtiss G (1993). Wisconsin card sorting test manual: Revised and expanded. Odessa, FL: Psychological Assessment Resources.
- Heaton RK, Grant I, & Matthews CG (1991). Comprehensive norms for an expanded Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications. Psychological Assessment Resources.
- Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR, & Conference Participants (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093–1129. 10.1080/13854040903155063 [PubMed: 19735055]
- Howe LLS (2012). Distinguishing genuine from malingered posttraumatic stress disorder in head injury litigation. In Reynolds CR & Horton AM Jr. (Eds.), *Detection of malingering during head injury litigation* (pp. 301–331). Springer Science + Business Media. 10.1007/978-1-4614-0442-2_11
- Jenkins MA, Langlais PJ, Delis D, & Cohen R (1998). Learning and memory in rape victims with posttraumatic stress disorder. *The American Journal of Psychiatry*, 155, 278–279. 10.1176/ajp.155.2.278 [PubMed: 9464211]
- Johnsen GE, Kanagaratnam P, & Asbjørnsen AE (2008). Memory impairments in posttraumatic stress disorder are related to depression. *Journal of Anxiety Disorders*, 22, 464–474. 10.1016/j.janxdis.2007.04.007 [PubMed: 17532601]
- Kang HK, Natelson BH, Mahan CM, Lee KY, & Murphy FM (2003). Posttraumatic stress disorder and chronic fatigue syndrome-like illness among Gulf War veterans: a population-based survey of 30,000 veterans. *American journal of epidemiology*, 157(2), 141–148. [PubMed: 12522021]
- Larrabee GJ (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410–425. 10.1076/clin.17.3.410.18089 [PubMed: 14704892]
- Larrabee GJ (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18, 625–630. 10.1017/S1355617712000240 [PubMed: 23057079]
- Lippa SM (2018). Performance validity testing in neuropsychology: a clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, 32, 391–421. 10.1080/13854046.2017.1406146 [PubMed: 29182444]
- Martin PK, Schroeder RW, & Odland AP (2015). Neuropsychologists' Validity Testing Beliefs and Practices: A Survey of North American Professionals. *The Clinical Neuropsychologist*, 29(6), 741–776. 10.1080/13854046.2015.1087597 [PubMed: 26390099]
- Martin PK, Schroeder RW, Olsen DH, Maloy H, Boettcher A, Ernst N, & Okut H (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, 34(1), 88–119. [PubMed: 31357918]
- Marx BP, Doron-Lamarca S, Proctor SP, & Vasterling JJ (2009). The influence of pre-deployment neurocognitive functioning on post-deployment PTSD symptom outcomes among Iraq-deployed Army soldiers. *Journal of the International Neuropsychological Society*, 15, 840–852. 10.1017/S1355617709990488 [PubMed: 19891817]
- Meyers JE, & Volbrecht ME (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, 18, 261–276. [PubMed: 14591459]
- Odland AP, Lammy AB, Martin PK, Grote CL, & Mittenberg W (2015). Advanced administration and interpretation of multiple validity tests. *Psychological Injury and Law*, 8, 46–63. 10.1007/s12207-015-9216-4
- Proto DA, Pastorek NJ, Miller BI, Romesser JM, Sim AH, & Linck JF (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, 614–624. [PubMed: 25252598]
- Reitan R, Wolfson D (1985) *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*. Tucson, AZ: Neuropsychology Press.
- Rey A (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

- Root JC, Robbins RN, Chang L, & Van Gorp WG (2006). Detection of inadequate effort on the California Verbal Learning Test-Second edition: Forced choice recognition and critical item analysis. *Journal of the International Neuropsychological Society*, 12, 688–696. 10.1017/S1355617706060838 [PubMed: 16961950]
- Rosen GM, & Powel JE (2003). Use of a Symptom Validity Test in the forensic assessment of Posttraumatic Stress Disorder. *Journal of Anxiety Disorders*, 17, 361–367. 10.1016/S0887-6185(02)00200-1 [PubMed: 12727128]
- Rosen GM, Sawchuk CN, Atkins DC, Brown M, Price JR, & Lees-Haley PR (2006). Risk of false positives when identifying malingered profiles using the trauma symptom inventory. *Journal of Personality Assessment*, 86, 329–333. 10.1207/s15327752jpa8603_08 [PubMed: 16740116]
- Schroeder RW, & Marshall PS (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist*, 25, 437–453. 10.1080/13854046.2011.556668 [PubMed: 21391153]
- Scott JC, Matt GE, Wrocklage KM, Crnich C, Jordan J, Southwick SM, Krystal JH, & Schweinsburg BC (2015). A quantitative meta-analysis of neurocognitive functioning in posttraumatic stress disorder. *Psychological Bulletin*, 141, 105–140. 10.1037/a0038039 [PubMed: 25365762]
- Sellbom M, Toomey JA, Wygant DB, Kucharski LT, & Duncan S (2010). Utility of the MMPI-2-RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment*, 22(1), 22. [PubMed: 20230148]
- Sellbom M, & Bagby RM (2010). Detection of overreported psychopathology with the MMPI-2 RF form validity scales. *Psychological Assessment*, 22(4), 757. [PubMed: 21133544]
- Sollman MJ, Ranseen JD, & Berry DT (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, 22, 325–335. 10.1037/a0018857 [PubMed: 20528060]
- Sullivan K, Kregel M, Proctor SP, Devine S, Heeren T, & White RF (2003). Cognitive functioning in treatment-seeking Gulf War veterans: pyridostigmine bromide use and PTSD. *Journal of Psychopathology and Behavioral Assessment*, 25, 95–103.
- Tanielian T, & Jaycox L (2008). *Invisible wounds of war: Psychological and cognitive injuries, their consequences, and services to assist recovery* (Vol. 1). Rand Corporation.
- Tombaugh TN (1996). *Test of memory malingering (TOMM)*. North Tonawanda, NY: Multi- Health Systems.
- Vasterling JJ, Brailey K, Constans JI, & Sutker PB (1998). Attention and memory dysfunction in posttraumatic stress disorder. *Neuropsychology*, 12, 125–133. 10.1037//0894-4105.12.1.125 [PubMed: 9460740]
- Vasterling JJ, Duke LM, Brailey K, Constans JI, Allain AN Jr, & Sutker PB (2002). Attention, learning, and memory performances and intellectual resources in Vietnam veterans: PTSD and no disorder comparisons. *Neuropsychology*, 16, 5–14. 10.1037//0894-4105.16.1.5 [PubMed: 11853357]
- Vasterling JJ, & Brailey K (2005). *Neuropsychological Findings in Adults with PTSD*. In Vasterling JJ & Brewin CR (Eds.), *Neuropsychology of PTSD: Biological, cognitive, and clinical perspectives* (pp. 178–207). The Guilford Press.
- Weathers FW, Huska JA, & Keane TM (1991). *PCL-C for DSM-IV*. Boston: National Center for PTSD-Behavioral Science Division.
- Wisdom NM, Pastorek NJ, Miller BI, Booth JE, Romesser JM, Linck JF, & Sim AH (2014). PTSD and cognitive functioning: Importance of including performance validity testing. *The Clinical Neuropsychologist*, 28(1), 128–145. [PubMed: 24354897]
- Wilkinson GS, & Robertson GJ (2006). *Wide range achievement test (WRAT4)*. Lutz, FL: Psychological Assessment Resources.
- Yehuda R, Golier JA, Tischler L, Stavitsky K, & Harvey PD (2005). Learning and Memory in Aging Combat Veterans with PTSD. *Journal of Clinical and Experimental Neuropsychology*, 27, 504–515. 10.1080/138033990520223 [PubMed: 15962694]
- Young JC, Sawyer RJ, Roper BL, & Baughman BC (2012). Expansion and re-examination of Digit Span effort indices on the WAIS-IV. *The Clinical Neuropsychologist*, 26(1), 147–159. [PubMed: 22268525]

Young JC, Roper BL, & Sawyer RJ (2011). Symptom validity test use and estimated base rates of failure in the VA Healthcare System. In Poster presented at the 9th annual meeting for the American Academy of Clinical Neuropsychology, Washington, DC.

Key Points

Question:

Does choice of performance validity criteria impact interpretation of cognitive results?

Findings: Results suggest that interpretation of neurocognitive data can be greatly influenced by the choice of performance validity test failure criteria applied.

Importance:

Findings highlight the importance of considering performance validity when interpreting cognitive data within clinical and research settings alike.

Next Steps:

Results warrant future investigation of PVT failure criteria in other conditions.

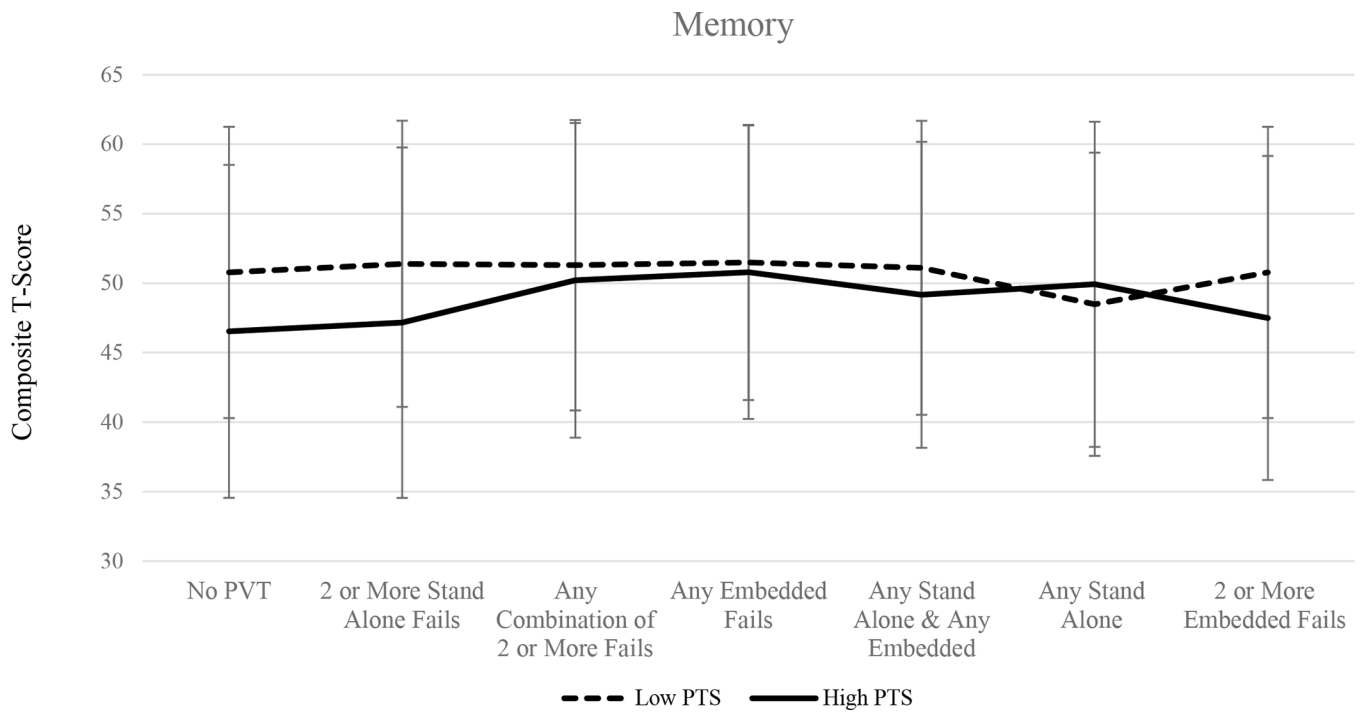


Figure 1. Mean memory composite T-scores among high PTS and low PTS groups across performance validity criteria.

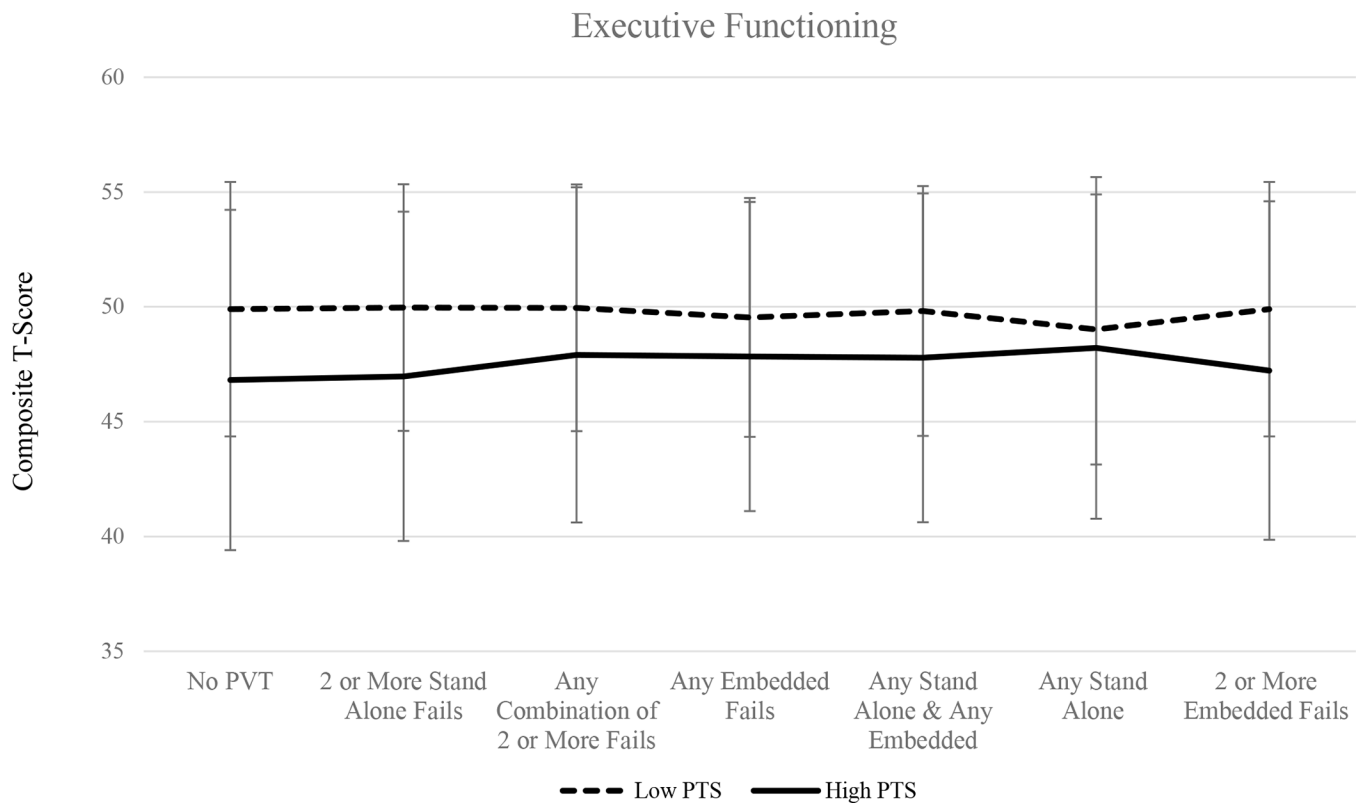


Figure 2. Mean executive functioning composite T-scores among high PTS and low PTS groups across performance validity criteria.

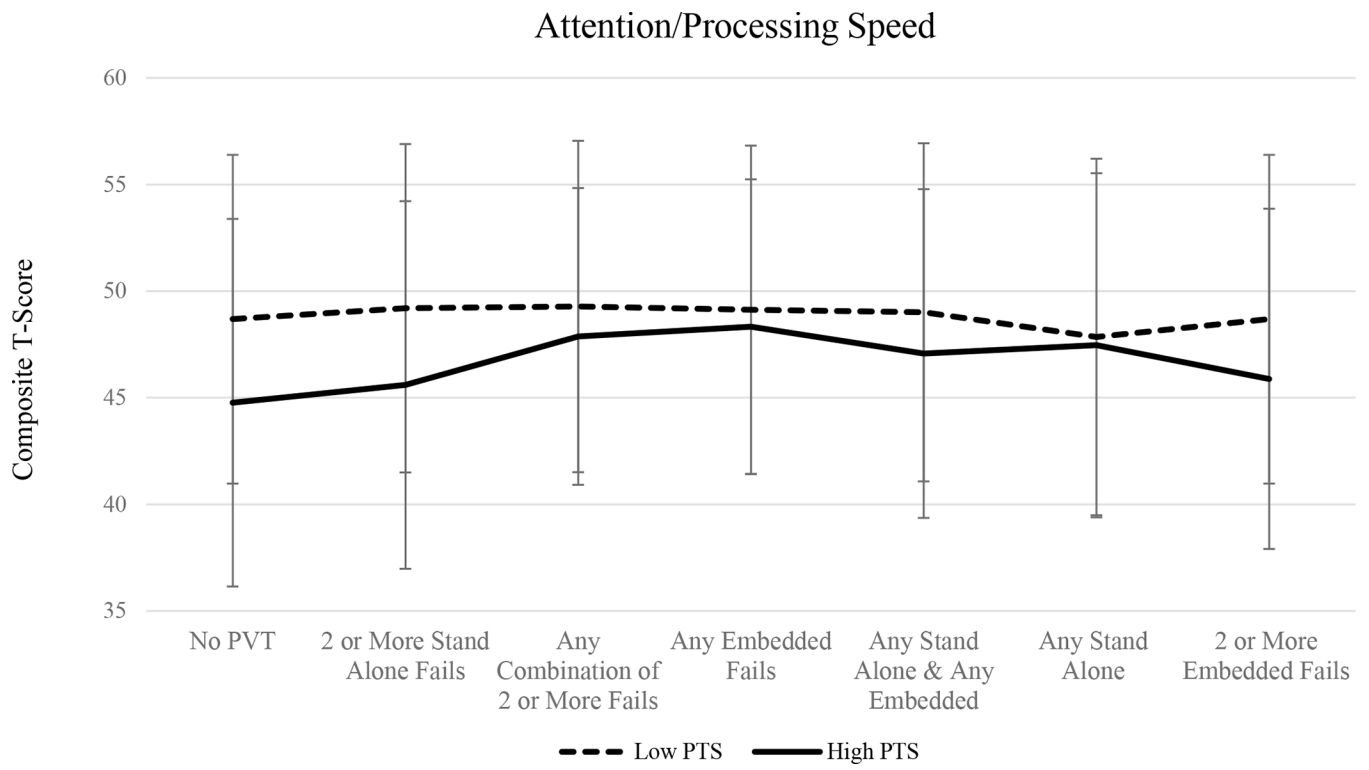


Figure 3. Mean attention/processing speed composite T-scores among high PTS and low PTS groups across performance validity criteria.

Table 1.

Participant demographic characteristic

Variable	Total Sample (N = 199)	High PTS (n = 140)	Low PTS (n = 59)	p
<i>Sociodemographic</i>				
Age (years)	31.9 (7.0)	31.3 (6.4)	33.5 (8.1)	0.04*
Sex (% women)	5.0	5.0	5.1	0.98
Marital Status (% married)	51.8	55.0	44.1	0.12
Education	13.1 (1.6)	13.0 (1.6)	13.3 (1.8)	0.24
WRAT-4 (Standard Score)	95.0 (13.8)	94.2 (15.0)	97.1 (10.4)	0.18
PCL-C (Raw Score)	57.5 (16.9)	66.4 (10.0)	36.4 (9.5)	
Employment Status (% employed)	42.2	40.7	45.8	0.51
Race (%)				0.09
White	68.8	65.0	78.0	
Black	11.6	14.3	5.1	
Hispanic	14.8	14.3	15.3	
Other	6.5	6.5	1.7	
Branch of Service (%)				0.73
Army	63.8	66.0	61.0	
Air Force	5.0	5.0	5.1	
Navy	7.0	7.9	5.1	
Marines	16.1	15.7	17.0	
National Guard	8.0	6.4	11.9	

Note: PCL-C= PTSD CheckList – Civilian Version PTSD= posttraumatic stress disorder; PTS = Post-Traumatic Stress; WRAT4=Wide Range Achievement Test 4

Table 2.

Performance validity test failure rates (%) across groups.

Measure	Total Sample	Low PTS	High PTS
Standalone PVTs (%)			
WMT	59.8	64	57.9
TOMM Trial 1	31.2	25	33.6
Rey 15	7.5	15.3	4.3
Embedded PVTs (%)			
CVLT-II (FC)	20.1	5.1	26.4
WAIS-IV (RDS)	13.1	3.4	17.1
WCST (FTMS)	9.5	11.9	8.6

Note: CVLT-II= California Verbal Learning Test-Second Edition (FC= Forced Choice); PTS= posttraumatic stress; TOMM= Test of Memory Malingering; WAIS-IV= Wechsler Adult Intelligence Scale -IV (RDS= Reliable Digit Span); WCST= Wisconsin Card Sorting Test (FTMS= Failure to Maintain Set)

Table 3.Cohen's *d* effect sizes by group across domain and failure criteria

Group	Sample Size	% Excluded by Group	Memory	Executive Functioning	Attention/Processing speed
No PVT Considered	N = 199		0.4402*	0.4746*	0.5609*
	Low PTS = 59				
	High PTS = 140				
Exclude 2 or More Embedded	N = 185	7%	0.3319*	0.4820*	0.3565*
	Low PTS = 59	0%			
	High PTS = 126	10%			
Exclude Any Embedded + Any Standalone	N = 151	24%	0.2328	0.3299*	0.3411*
	Low PTS = 52	12%			
	High PTS = 99	29%			
Exclude 2 or More Standalone	N = 134	33%	0.3869*	0.4725*	0.4909*
	Low PTS = 42	29%			
	High PTS = 92	34%			
Exclude Any Embedded	N = 131	34%	0.1699	0.2814	0.2021
	Low PTS = 47	20%			
	High PTS = 84	40%			
Exclude Any Combo of 2 or More	N = 110	45%	0.1285	0.3277	0.3223
	Low PTS = 40	32%			
	High PTS = 70	50%			
Exclude Any Standalone	N = 78	61%	0.0802	0.1196	0.1956
	Low PTS = 20	66%			
	High PTS = 58	59%			

* Indicates significant group differences ($p < .05$) within domain PTS = posttraumatic stress, PVT = performance validity test