Comprehensive Analysis of MMPI-2-RF Symptom Validity Scales and Performance Validity Test Relationships in a Diverse Mixed Neuropsychiatric Setting

Adam B. De Boer1 · Matthew S. Phillips1 · Kearston C. Barwegen1 · Maximillian A. Obolsky1 · Andrew A. Rauch1 · Stephen D. Pesanti1 · Phoebe Ka Yin Tse1 · Gabriel P. Ovsiew1 · Kyle J. Jennette1 · Zachary J. Resch1 · Jason R. Soble1,2

Received: 14 June 2022 / Accepted: 24 October 2022 / Published online: 3 November 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
The utility of symptom (SVT) and performance (PVT) validity tests has been independently established in neuropsychological evaluations, yet research on the relationship between these two types of validity indices is limited to circumscribed populations and measures. This study examined the relationship between SVTs on the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) and PVTs in a mixed neuropsychiatric setting. This cross-sectional study included data from 181 diagnostically and demographically diverse patients with neuropsychiatric conditions referred for outpatient clinical neuropsychological evaluation at an academic medical center. All patients were administered a uniform neuropsychological battery, including the MMPI-2-RF and five PVTs (i.e., Dot Counting Test; Medical Symptom Validity Test; Reliable Digit Span; Test of Memory Malingering-Trial 1; Word Choice Test). Nonsignificant associations emerged between SVT and PVT performance. Although the Response Bias Scale was most predictive of PVT performance, MMPI-2-RF SVTs generally had low classification accuracy for predicting PVT performance. Neuropsychological test performance was related to MMPI-2-RF SVT status only when overreporting elevations were at extreme scores. The current study further supports that SVTs and PVTs measure unique and dissociable constructs among diverse patients with neuropsychiatric conditions, consistent with literature from other clinical contexts. Therefore, objective evidence of symptom overreporting on MMPI-2-RF SVTs cannot be interpreted as definitively indicating invalid performance on tests of neurocognitive abilities. As such, clinicians should include both SVTs and PVTs as part of a comprehensive neuropsychological evaluation as they provide unique information regarding performance and symptom validity.

Keywords Symptom validity test · Performance validity test · MMPI-2-RF · Neuropsychology · Assessment

Introduction
Objective assessment of the validity of examinees’ symptom reporting and cognitive test performance is an essential component of psychological and neuropsychological evaluations to ensure the accuracy and credibility of examinees’ responses/reported symptoms on self-report inventories and objective neuropsychological test performance. Accordingly, practice standards have evolved and now specifically require the routine, objective assessment of performance and symptom validity across all evaluations (American Academy of Clinical Neuropsychology, 2007; Bush et al., 2005; Sweet et al., 2021). To meet this practice standard, neuropsychological test batteries are often considerably lengthened in order to include measures of performance and symptom validity because they represent conceptually and empirically distinct, albeit variably overlapping, constructs that need to be assessed independently (Larrabee, 2012).

Performance validity tests (PVTs) objectively evaluate the degree to which examinees’ cognitive test performance is an accurate reflection of their true abilities rather than feigned impairment or suboptimal test engagement (Larrabee, 2012; Soble et al., 2017; Van Dyke et al., 2013). Current practice standards call for the use of a battery of PVTs that assess
various cognitive domains (e.g., memory, attention) interspersed throughout the entirety of the evaluation to ensure continuous sampling of performance validity (Boone, 2009). A recent review by Soble et al. (2021b) reaffirms that the development and cross-validation of PVTs has grown exponentially over the past two decades. This proliferation was largely driven by the development and expansion of embedded validity indicators within well-validated tests of cognitive ability. By contrast, the primary aim of symptom validity tests (SVTs) is to differentiate patients with credible symptom reporting from those who may be exaggerating or dishonestly representing complaints on psychological and symptom inventories (Fokas & Brovko, 2020). Overall, there are far fewer SVTs relative to PVTs because most routine self-report psychological symptom measures do not contain SVTs (e.g., Beck Depression Inventory [Beck et al., 1996]), although some exceptions exist (e.g., Clinical Assessment of Attention Deficit [Bracken & Boatwright, 2005]); Neurobehavioral Symptom Inventory [Soble et al., 2014; Vanderploeg et al., 2014]). Moreover, most available SVTs are contained in lengthier and more broadband measures of psychopathology and personality, such as the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) and Personality Assessment Inventory (PAI; Morey, 1991), thereby increasing battery length, examination costs, and burden on the examinee.

To date, prior studies examining general relationships between SVT and PVT performance have yielded equivocal findings. Similarly, some recent studies have also demonstrated that symptom and performance validity represent distinct constructs in the context of adult ADHD evaluations (e.g., Leib et al., 2021; White et al., 2022) and neuropsychological evaluations in Veteran Affairs (VA) settings (e.g., Bomaye et al., 2020; Ingram et al., 2019, 2020; Ord et al., 2021). For instance, Aase et al. (2021) and Shura et al. (2021) found that PVTs and SVTs were largely dissociable among post-deployment veterans with conditions including posttraumatic stress disorder (PTSD) and concussion/mild traumatic brain injury (TBI). However, not all research has supported the notion that they are independent constructs as more robust relationships between MMPI-2-RF SVTs and PVTs have been shown in medicolegal and forensic populations (e.g., Gervais et al., 2007, 2011). Within the context of neuropsychological evaluations, the MMPI-2-RF validity scales have arguably received the most empirical attention of any SVT, likely due to their breadth in detecting symptom magnification or feigning (Ingram & Ternes, 2016), covering psychiatric, somatic, and cognitive symptoms, particularly in medicolegal or disability settings (e.g., Gervais et al., 2007, 2011). That said, studies assessing the relationship between MMPI-2-RF SVTs and PVT performance in general clinical neuropsychiatric samples are limited. Considering prior equivocal findings regarding general relationships between SVTs and PVTs across non-medicolegal clinical populations, further exploration of the relationship (or lack thereof) among the MMPI-2-RF validity scales and PVTs within the context of clinical neuropsychological evaluations is needed. As such, the primary objective of this study was to provide a detailed examination of the relationship between the MMPI-2-RF validity scales and performance across a battery of well-validated freestanding and embedded PVTs in an ethnoracially and diagnostically diverse mixed neuropsychiatric sample.

**Method**

**Participants**

This cross-sectional study analyzed data from a large mixed neuropsychiatric sample of 277 patients referred for neuropsychological evaluation at an urban academic medical center between 2018 and 2022. Evaluations were completed for the purposes of differential diagnosis, characterization of cognitive status, treatment planning, and/or pre-neurosurgical baselining. All patients consented to collecting their test scores as part of a larger, IRB-approved neuropsychological database protocol. Inclusion criteria included (1) the examinee was administered and completed the MMPI-2-RF and the five criterion PVTs included in the reference standard (see below for more details) and (2) no evidence of MMPI-2-RF Variable Response Inconsistency Scale (VRIN-r) or True Response Inconsistency Scale (TRIN-r) elevations (e.g., T-score ≥ 80), indicating excessive non-content-based responding. Two patients were missing one criterion PVT; 77 were not administered an MMPI-2-RF, frequently due to significant cognitive/neurobehavioral disturbance preventing them from tolerating the protocol; and 17 patients had MMPI-2-RF VRIN-r or TRIN-r elevations ≥ 80 T. The data for these 96 patients was excluded from all subsequent analysis, resulting in a final sample size of 181 cases. All examinees within the final sample completed every test within the standardized neuropsychological test battery described below in the “Measures” section. Tables 1 and 2 report sample demographic and diagnostic data, respectively.

**Measures**

**Performance Validity Tests** All patients were administered the following five freestanding and embedded PVTs which comprise the reference standard during the course of their neuropsychological evaluations: Dot Counting Test (DCT; Boone et al., 2002; failure rate: 14.4%), Medical Symptom Validity Test (MSVT; Green, 2004; Resch et al., 2022a; failure rate: 19.7%), Reliable Digit Span (RDS; Schroeder et al., 2012; failure rate: 9.9%), Test of Memory Malingering

© Springer
(TOMM) Trial 1 (Martin et al., 2020; failure rate: 22.7%), and the Word Choice Test (WCT; Bernstein et al., 2021; Neale et al., 2022; failure rate: 13.3%). Patients with two or more PVT failures among the reference standard were classified as having invalid neuropsychological test performance, in line with current practice standards and empirically supported methodological approaches in PVT research (e.g., Jennette et al., 2021; Larrabee, 2008; Rhoads et al., 2021; Sweet et al., 2021). Among the final sample, 146 examinees (81%) were classified as having demonstrated valid neuropsychological test performance based on the reference standard, whereas 35 (19%) were classified as having invalid performance.

**Minnesota Multiphasic Personality Inventory-2-Restructured Form (Ben-Porath & Tellegen, 2008)** The MMPI-2-RF is a 338-item standardized psychometric test of adult personality and psychopathology, containing a total of 51 scales, nine of which assess symptom validity. It is used to aid clinicians in the assessment of psychiatric disorders, the identification of specific problem areas, and treatment planning. The MMPI-2-RF was validated using a gender-balanced normative sample drawn from the MMPI-2 norms consisting of 2,276 men and women between the ages of 18 and 80 and includes comparison groups from inpatient, outpatient, and forensic settings. Of interest to the current study are the nine MMPI-2-RF validity scales, two of which assess for non-content-based responding (i.e., VRIN-r; TRIN-r); five of which assess for overreporting of infrequent symptoms, rare psychiatric symptoms, and unusual somatic and cognitive complaints (i.e., F-r; Fp-r; Fs; FBS-r; RBS); and two of which assess defensiveness/symptom underreporting (i.e., L-r; K-r). The five overreporting scales were operationalized based on standard interpretation guidelines (Ben-Porath, 2012) with F-r, Fp-r, Fs, FBS-r, and/or RBS ≥ 100 T indicating definite overreporting and F-r 79-99 T, Fp-r 70-99 T and Fs, FBS-r, and/or RBS 80-99 T denoting possible overreporting.

**Neuropsychological Test Battery** All patients completed a core neuropsychological test battery for comprehensive assessment of neurocognitive status. This battery included the Test of Premorbid Function (TOPF; Pearson, 2009), Verbal Fluency (F/A/S and Animal Naming; Heaton et al., 2004), Rey Auditory Verbal Learning Test (RAVLT; Schmidt, 1996), Brief Visuospatial Memory Test-Revised (BVMT-R; Benedict, 1997), Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; Wechsler, 2008) Processing Speed Intact (PSI), Trail Making Test (TMT; Heaton et al., 2004), and Stroop Color and Word Test (Golden, 1978). Of note,
although embedded PVTs derived from the RAVLT (Boone et al., 2005; Pliskin et al., 2021; Soble et al., 2021a) and BVMT-R (Bailey et al., 2018; Resch et al., 2022b), Verbal Fluency, TMT, and Stroop (Khan et al., 2022; White et al., 2020) have been identified; these embedded indicators were not included in the reference standard in order to avoid criterion contamination by keeping the neurocognitive tests fully independent from the criterion PVTs used to establish the validity groups. Among the 30 total patients who were actively compensation-seeking at the time of evaluation (see Table 1), 27% (8/30) obtained ≥ 2 PVT failures and were classified into the invalid group, whereas the remaining 73% (22/30) of those who were actively compensation-seeking demonstrated valid test performance (i.e., ≤ 1 PVT failure). Among the remaining 151 who were not actively compensation-seeking, 18% (27/151) had invalid PVT performance and 82% (124/151) had valid performance. There were nonsignificant differences in validity group membership (i.e., valid/invalid) based on compensation-seeking status, $X^2(1,181) = 1.24, p = 0.266$, indicating the presence of financial incentive did not meaningfully influence validity status.

### Statistical Analyses

Elevation base rates for the seven MMPI-2-RF symptom overreporting and underreporting validity scales (i.e., F-r, FP-r, Fs, FBS-r, RBS, L-r, and K-r) were calculated using the test developer’s recommended elevation thresholds for possible and definite over/underreporting (Ben-Porath, 2012). Spearman correlations between the MMPI-2-RF validity scales and the criterion PVTs were calculated, for the overall sample and then separately for the subsamples with valid and invalid performance to examine for differences in PVT/SVT associations between performance validity subgroups. Analyses of variance (ANOVAs) tested for differences on the MMPI-2-RF validity scale scores between those with valid and invalid neuropsychological performance, and chi-square tests examined for differences in elevation rates (i.e., elevated/unelevated) between validity groups. Receiver operator characteristic (ROC) curve analyses examined the ability of each MMPI-2-RF validity scale to differentiate those with valid and invalid neuropsychological performance, and chi-square tests examined for differences in elevation rates (i.e., elevated/unelevated) between validity groups. Receiver operator characteristic (ROC) curve analyses examined the ability of each MMPI-2-RF validity scale to differentiate those with valid and invalid neuropsychological test performance. For ROC analyses with a significant area under the curve (AUC), optimal cut-scores that maximized sensitivity while maintaining acceptable specificity (i.e., 90%; Boone, 2013) were identified. Finally, ANOVAs examined for differences in neurocognitive test performance based on the presence/absence and number of MMPI-2-RF overreporting validity scale elevations. The false discovery rate (FDR) procedure with a 0.05 maximum FDR was implemented to control the familywise error rate associated with multiple ANOVA comparisons (Benjamini & Hochberg, 1995).

### Results

Elevation base rates for the seven MMPI-2-RF symptom reporting validity scales for the overall sample are presented in Table 3. In brief, F-r and RBS scale elevations were most common, whereas K-r had the lowest elevation rates. A similar pattern emerged when elevation rates across these seven validity scales were further subdivided by possible and definite over/underreporting (Table 4). Finally, the total number of elevations across the five overreporting scales (F-r, FP-r, Fs, FBS-r, and RBS) is presented in Table 5. In total, 43% of patients had no overreporting scale elevations, whereas 38% had two or more scale elevations.
As seen in the upper section of Table 6, correlations between the MMPI-2-RF validity scales and the criterion PVTs revealed nonsignificant to small effects for the overall sample. In the subsample with valid neuropsychological test performance (middle section of Table 6), correlations were generally nonsignificant. Similarly, correlations among the invalid group (lower section of Table 6) also were almost entirely nonsignificant save for a small significant association between two MSVT indices and TOMM T1 and RBS. Together, these results indicate MMPI-2-RF validity scales were largely independent from PVT performance.

As noted in Table 7, ANOVAs revealed those with valid PVT performance had significantly fewer scale elevations and lower mean scores on the MMPI-2-RF validity scales than those with invalid PVT performance, except for the K-r scale, with small to medium effects. F-r and RBS yielded the most significant differences and largest effects between groups. When considering MMPI-2-RF higher order and Restructured Clinical (RC) scales (Table 8), those with invalid PVT performance scored significantly higher on the RC2 and RC8 scales, although effects were small.

The MMPI-2-RF validity scales significantly differentiated valid and invalid PVT performance for the F-r, Fs, FBS-r, and RBS scales, in addition to the total number of overreporting scale elevations (Table 9). However, AUC values revealed low classification accuracy, ranging from 0.612 (Fs) to 0.690 (RBS). Sensitivity values ranged from 22.9% (FBS) to 40.0% (RBS) with specificity values approximating 90% at their respective optimal cut-scores. Of note, the RBS scale yielded the highest AUC, sensitivity, and specificity values of all the scales/total scales, indicating that it had the most robust classification accuracy. As was expected, optimal cut-scores were above the threshold for possible overreporting, but below the definite overreporting threshold based on the test-publisher recommendations (Ben-Porath, 2012).

As seen in Table 10, when the sample was trichotomized by 0, 1, or ≥ 2 possible overreporting elevations, cognitive test scores were not significantly different across groups. In contrast, more significant differences emerged when examining definite overreporting elevations (Table 11). Namely, WAIS-IV PSI and TMT Part A performances were significantly different between those with 0 and ≥ 2 overreporting elevations. Significant differences between those with 0 and 1 overreporting elevations were discovered for Rey Auditory Verbal Learning Test (RAVLT) learning and delayed recall scores, but not between those with 0 and ≥ 2 elevations. Due to small group sizes when the sample was trichotomized by definite overreporting, 1 and ≥ 2 overreporting groups were collapsed into a single group (i.e., ≥ 1 definite overreporting elevations) for additional analysis, allowing for comparison.

---

**Table 3** MMPI-2-RF Validity Scale elevations

| Elevated scale | n | % |
|----------------|---|---|
| F-r           | 63 | 34.8 |
| Fp-r          | 37 | 20.4 |
| Fs            | 48 | 26.5 |
| FBS-r         | 40 | 22.1 |
| RBS           | 66 | 36.5 |
| L-r           | 48 | 26.5 |
| K-r           | 20 | 11.0 |

Cutoff values are as follows: F-r ≥ 79; Fp-r ≥ 70; Fs ≥ 80; FBS-r ≥ 80; RBS ≥ 80; L-r ≥ 64; K-r ≥ 59

**Table 4** MMPI-2-RF Validity Scale elevations stratified

| Scale | No elevation | Possible over/underreporting | Definite over/underreporting |
|-------|--------------|------------------------------|-----------------------------|
|       | n            | %   | n            | %   | n            | %   |
| F-r   | 118 (< 79 T) | 65.2 | 43 (79–99 T) | 23.8 | 20 (≥ 100 T) | 11.0 |
| Fp-r  | 144 (< 70 T) | 79.6 | 28 (70–99 T) | 15.5 | 9 (> 100 T)  | 5.0  |
| Fs    | 133 (< 80 T) | 73.5 | 36 (80–99 T) | 19.9 | 12 (≥ 100 T) | 6.6  |
| FBS-r | 141 (< 80 T) | 77.9 | 37 (80–99 T) | 20.4 | 3 (≥ 100 T)  | 1.7  |
| RBS   | 115 (< 80 T) | 63.5 | 51 (80–99 T) | 28.2 | 15 (≥ 100 T) | 8.3  |
| L-r   | 133 (< 64 T) | 73.5 | 40 (65–79 T) | 22.1 | 8 (≥ 80 T)  | 4.4  |
| K-r   | 161 (< 59 T) | 89.0 | 19 (60–69 T) | 10.5 | 1 (≥ 70 T)  | 0.6  |

---
of those with no definite overreporting elevations to those with any elevations (upper section of Table 11). Similar to previous group analysis, WAIS-IV PSI, TMT Part A, and RAVLT learning and delayed recall scores were all significantly higher (i.e., better performance) for those with no elevations than those in the overreporting group, although effects were generally small. F/A/S and BVMT-R scores were also significantly different between groups, again with small effects.

**Discussion**

This study evaluated the relationship between MMPI-2-RF SVTs and PVT performance in a mixed neuropsychiatric sample at a large Midwestern academic medical center. Overall, PVT performance had nonsignificant to weak associations with symptom reporting on the MMPI-2-RF validity scales, and this general lack of significant relationships held constant among both the valid and invalid subsamples. Moreover, among the entire sample, elevations for the MMPI-2-RF F-r and RBS scales were the most common among those with both possible and definite overreporting. F-r, Fs, FBS-r, and RBS scales were able to significantly discriminate valid from invalid PVT performance, albeit with relatively low classification accuracies (0.612–0.690) and generally low sensitivity, ranging from 23 to 40%. Optimal cut-scores were consistently higher than the minimum score thresholds for possible overreporting but below test-publisher recommendations for definite overreporting minimum thresholds (Ben-Porath, 2012). When comparing cognitive test scores to the number of possible overreporting elevations (i.e., 0, 1, or ≥ 2), few significant differences were identified, with negligible effects. When MMPI-2-RF validity elevations were operationalized using definite overreporting thresholds rather than possible overreporting thresholds, more statistically significant cognitive differences emerged, revealing lower scores on WAIS-IV PSI, TMT Part A, and RAVLT learning and delayed recall measures in those with elevated validity.
scale scores; however, effects were small and scores generally fell within the same clinical interpretive range (e.g., low average to average range) across the SVT groups.

The study findings further support previous literature indicating that SVTs and PVTs are dissociable, thereby highlighting the benefits of independently assessing each construct within a neuropsychological evaluation (Bomyea et al., 2020). To our knowledge, no previous study has explored these relationships within a mixed neuropsychiatric sample with PVTs derived from (or appearing to be derived from) diverse cognitive modalities. By addressing this gap in the literature, these findings can be generalized to populations beyond the VA and forensic/medicolegal contexts where most of this research has been conducted (e.g., Copeland et al., 2015; Gervais et al., 2007; Ord et al., 2021; Van Dyke et al., 2013; Whitney et al., 2008). Additionally, results were consistent with previous research showing that performance on SVTs and PVTs provide nonredundant validity information, even within diverse populations. Cognitive performance was similar across SVT validity groups in this sample, only yielding small effects on cognitive scores when MMPI-2-RF validity elevations were operationalized as definite overreporting, meaning that unequivocally elevated MMPI-2-RF validity scores were necessary before there was a clear impact on cognitive test scores. That said, most MMPI-2-RF elevations fell within the possible overreporting range, indicating the dissociability of SVTs and PVTs remains for most cases and comes into question only when symptom exaggeration is extreme.

Table 7 Comparing MMPI-2-RF validity scale scores by performance validity status

|                  | PVTs Valid (n = 146) | PVTs Invalid (n = 35) | χ² | F    | ηp²  |
|------------------|----------------------|-----------------------|----|------|------|
|                  | M (SD)              | N   | %   | M (SD) | N   | %   |
| F-r              | 69.73 (18.39)       | 103 | 70.5 | 82.91 (22.83) | 15 | 42.9 | 13.147*** | .068 |
| Not elevated     | 43                  | 29.4 | 20 | 57.1 |
| Elevated scale   | 118                 | 80.8 | 26 | 74.3 | 0.742 |
| Fp-r             | 59.67 (16.41)       | 66.49 (20.30) | 4.424* | .024 |
| Not elevated     | 28                  | 19.2 | 9 | 25.7 |
| Elevated scale   | 112                 | 76.7 | 21 | 60.0 | 4.047* |
| Fs               | 66.21 (18.16)       | 75.97 (23.86) | 7.168** | .039 |
| Not elevated     | 34                  | 23.3 | 14 | 40.0 |
| Elevated scale   | 112                 | 76.7 | 21 | 60.0 | 4.047* |
| FBS-r            | 66.72 (12.94)       | 74.20 (15.16) | 8.818** | .047 |
| Not elevated     | 119                 | 81.5 | 22 | 62.9 |
| Elevated scale   | 27                  | 18.5 | 22 | 62.9 |
| RDd              | 69.74 (16.24)       | 83.86 (20.70) | 19.065*** | .096 |
| Not elevated     | 99                  | 67.8 | 16 | 45.7 | 5.948* |
| Elevated scale   | 47                  | 32.2 | 19 | 54.3 |
| RC1              | 57.18 (11.74)       | 58.97 (9.99) | .691 | .004 |
| Not elevated     | 109                 | 74.7 | 24 | 68.6 |
| Elevated scale   | 37                  | 25.3 | 11 | 31.4 |
| RC2              | 46.32 (9.87)        | 44.14 (10.22) | 1.358 | .008 |
| Not elevated     | 130                 | 89.0 | 31 | 88.6 | 0.006 |
| Elevated scale   | 16                  | 11.0 | 4 | 11.4 |

All p-values are false discovery rate (FDR)-corrected p-values

*p < .05; ** p < .01; ***p < .001

Table 8 Comparing MMPI-2-RF higher order and RC Scale scores by performance validity status

|                  | PVTs Valid (n = 146) | PVTs Invalid (n = 35) | F   | ηp²  |
|------------------|----------------------|-----------------------|-----|------|
|                  | M (SD)              | M (SD)               |     |     |
| EID              | 59.00 (14.00)       | 63.89 (14.61)        | 3.381 | .019 |
| THD              | 55.98 (12.03)       | 61.66 (17.52)        | 5.187 | .028 |
| BXD              | 50.83 (10.89)       | 50.03 (10.68)        | 0.153 | .001 |
| RDd              | 60.90 (12.72)       | 65.14 (11.95)        | 3.209 | .018 |
| RC1              | 67.25 (12.61)       | 72.49 (16.10)        | 4.353 | .024 |
| RC2              | 59.81 (13.70)       | 67.20 (16.99)        | 7.461* | .040 |
| RC3              | 51.67 (10.28)       | 56.14 (11.75)        | 5.049 | .027 |
| RC4              | 53.58 (10.52)       | 53.29 (10.70)        | 0.022 | .000 |
| RC6              | 57.88 (12.75)       | 61.40 (17.05)        | 0.173 | .010 |
| RC7              | 55.45 (13.52)       | 59.46 (13.50)        | 2.487 | .014 |
| RC8              | 57.48 (12.33)       | 64.80 (14.41)        | 9.301* | .049 |
| RC9              | 47.92 (9.95)        | 47.00 (11.76)        | 0.223 | .001 |

All p-values are false discovery rate (FDR)-corrected p-values

*p < .05
Despite their dissociable nature, the MMPI-2-RF symptom validity scales were shown to, with variable accuracy, function similarly to a PVT by detecting invalid cognitive performance, particularly with the RBS scale. This finding was not unexpected considering the RBS scale was specifically designed to detect cognitive response bias, effectively differentiating between those who did and did not pass several memory-based PVTs (Gervais et al., 2007). Yet, consistent with extant literature, classification accuracy was relatively low (i.e., AUCs all under 0.70), indicating MMPI-2-RF symptom validity scales should not be independently used to evaluate performance validity in neuropsychological evaluations. SVT elevations rates were also higher among those with invalid PVT performance, although a notable caveat is that MMPI-2-RF elevations are common even among validly performing patients, particularly

| Table 9 | Predicting performance validity status by MMPI-2-RF validity scales |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | AUC | Cut-score | SN | SP | 10% Base Rate | PPV | NPV | 20% Base Rate | PPV | NPV | 30% Base Rate | PPV | NPV | 40% Base Rate | PPV | NPV |
| F-r | .671** | ≥ 90 | .429 | .863 | 0.26 | 0.93 | 0.44 | 0.86 | 0.57 | 0.87 | 0.68 | 0.69 |
| | | ≥ 95 | .314 | .884 | 0.23 | 0.93 | 0.40 | 0.84 | 0.54 | 0.75 | 0.64 | 0.66 |
| | | ≥ 99 | .257 | .925 | 0.28 | 0.93 | 0.46 | 0.83 | 0.59 | 0.74 | 0.70 | 0.65 |
| Fp-r | .601 | - | - | - | - | - | - | - | - |
| Fs | .612* | ≥ 79 | .400 | .767 | 0.16 | 0.92 | 0.30 | 0.84 | 0.42 | 0.75 | 0.53 | 0.66 |
| | | ≥ 87 | .314 | .884 | 0.23 | 0.92 | 0.40 | 0.84 | 0.54 | 0.75 | 0.64 | 0.66 |
| | | ≥ 95 | .286 | .918 | 0.28 | 0.92 | 0.47 | 0.84 | 0.60 | 0.75 | 0.70 | 0.66 |
| FBS-r | .638* | ≥ 82 | .314 | .863 | 0.20 | 0.92 | 0.36 | 0.83 | 0.50 | 0.75 | 0.60 | 0.65 |
| | | ≥ 85 | .229 | .890 | 0.19 | 0.91 | 0.34 | 0.83 | 0.47 | 0.73 | 0.58 | 0.63 |
| | | ≥ 88 | .171 | .918 | 0.19 | 0.91 | 0.34 | 0.82 | 0.47 | 0.72 | 0.58 | 0.62 |
| RBS | .690*** | ≥ 86 | .486 | .863 | 0.28 | 0.91 | 0.47 | 0.87 | 0.60 | 0.80 | 0.70 | 0.72 |
| | | ≥ 90 | .400 | .918 | 0.35 | 0.91 | 0.55 | 0.86 | 0.68 | 0.78 | 0.76 | 0.70 |
| | | ≥ 95 | .343 | .945 | 0.41 | 0.93 | 0.61 | 0.85 | 0.73 | 0.77 | 0.81 | 0.68 |
| L-r | .558 | - | - | - | - | - | - | - | - |
| K-r | .434 | - | - | - | - | - | - | - | - |
| Total overreporting | .646** | ≥ 3 | .457 | .788 | 0.19 | 0.93 | 0.35 | 0.85 | 0.48 | 0.77 | 0.59 | 0.69 |
| | | ≥ 4 | .286 | .904 | 0.25 | 0.92 | 0.43 | 0.84 | 0.56 | 0.75 | 0.67 | 0.66 |
| | | ≥ 5 | .114 | .945 | 0.19 | 0.91 | 0.34 | 0.81 | 0.47 | 0.71 | 0.58 | 0.62 |

Bolded scores refer to optimal cut-scores

AUC area under the curve, PPV positive predictive value, NPV negative predictive value

*p < .05; **p < .01; ***p < .001

Table 10 Comparing neurocognitive test performance by scores on the MMPI-2-RF overreporting validity scales

| RAVLT | BVMT-R | WAIS-IV PSI |
|-------|-------|-----------|
| F/A/S | Animals | RAVLT trials 1–5 |
| M (SD) | M (SD) | M (SD) |
| 46.23 (11.86) | 45.09 (11.68) | 41.21 (14.03) |
| 46.23 (11.40) | 43.86 (10.84) | 40.71 (11.81) |
| 42.97 (10.45) | 44.47 (10.71) | 37.24 (10.69) |
| 1.849 | 0.157 | 2.018 |
| 0.20 | 0.02 | 0.02 |
| RAVLT delayed recall | BVMT-R trials 1–3 | BVMT-R delayed recall |
| M (SD) | M (SD) | M (SD) |
| 42.49 (13.85) | 43.67 (16.36) | 44.24 (17.73) |
| 41.26 (11.94) | 43.17 (13.75) | 45.80 (13.24) |
| 39.28 (12.72) | 42.66 (15.02) | 42.69 (16.19) |
| 1.098 | 0.078 | 0.437 |
| 0.12 | 0.01 | 0.05 |
| BVMT-R delayed recall | WAIS-IV PSI | Trail Making Test A |
| M (SD) | M (SD) | M (SD) |
| 44.24 (17.73) | 95.44 (15.62) | 49.59 (12.36) |
| 45.80 (13.24) | 97.43 (16.54) | 51.57 (13.35) |
| 42.69 (16.19) | 91.16 (15.57) | 44.84 (12.95) |
| 2.232 | 4.035 | 0.870 |
| 0.24 | 0.04 | 0.10 |
| Trail Making Test B | Stroop Color-Word |
| M (SD) | M (SD) |
| 44.23 (13.44) | 45.71 (11.62) |
| 46.63 (12.22) | 46.43 (11.31) |
| 43.09 (12.61) | 44.15 (9.44) |
| 0.633 | .033 |
| .07 | .007 |

RAVLT Rey Auditory Verbal Learning Test, BVMT-R Brief Visuospatial Memory Test-Revised, WAIS-IV PSI Wechsler Adult Intelligence Scale-Fourth Edition processing speed index

 Springer
for overreporting scales (Ingram et al., 2020). Part of this observed dissociation between SVTs and PVTs is that unlike PVTs, where a generally accepted threshold for invalidity has been established in the literature (i.e., ≥ 2 PVT fails; Larrabee, 2008; Rhoads et al., 2021), clearly delineated benchmarks have yet to be established for SVTs. Using the MMPI-2-RF as an example, it is generally accepted that elevations in the definite overreporting ranges (see Ben-Porath, 2012) reflect invalid symptom reporting; however, for the possible symptom overreporting range, there is no clear, empirically derived benchmark for the optimal number of scale elevations needed to confidently conclude symptom invalidity have been established. Further complicating matters is that certain clinical populations with independently and objectively verified valid symptom reporting have been shown to produce mild elevations on some MMPI-2-RF validity scales relative to the normative sample (e.g., electrical injury: Soble et al., 2019).

This study contained several noteworthy methodological strengths. First, it consisted of a large and demographically diverse sample that is representative of large urban medical centers. Secondly, five well-established and widely used criterion PVTs were used as a reference standard to establish validly and invalidly performing subgroups within the sample. Many extant studies have been conducted with reference standards consisting of only one or two criterion PVTs, which likely results in inaccurate validity classifications due to overreliance on limited validity information.

### Table 11 Comparing neurocognitive test performance by MMPI-2-RF overreporting validity scales (definite overreporting)

|                | 0 Definite overreporting elevations (n = 148) | ≥ 1 Definite overreporting elevation (n = 33) | F    | η²  |
|----------------|---------------------------------------------|---------------------------------------------|------|-----|
|                | M (SD)                                      | M (SD)                                      |      |     |
| F/A/S          | 45.97 (11.18)                               | 40.88 (11.21)                               | 5.614* | .030 |
| Animals        | 45.04 (10.82)                               | 42.73 (12.34)                               | 1.171 | .006 |
| RAVLT trials 1–5 | 40.91 (12.67)                               | 33.85 (10.11)                               | 8.966* | .048 |
| RAVLT delayed recall | 42.36 (13.06)                               | 35.15 (11.67)                               | 8.523* | .045 |
| BVMT-R trials 1–3  | 44.57 (15.20)                               | 37.03 (14.48)                               | 6.751* | .036 |
| BVMT-R delayed recall | 45.07 (15.93)                               | 39.00 (17.38)                               | 3.788 | .021 |
| WAIS-IV PSI    | 95.90 (16.12)                               | 86.67 (12.43)                               | 9.544* | .051 |
| Trail Making Test A | 49.50 (12.46)                               | 42.30 (13.85)                               | 8.636* | .046 |
| Trail Making Test B | 44.80 (12.62)                               | 41.85 (14.02)                               | 1.421 | .008 |
| Stroop Color-Word | 45.36 (10.86)                               | 44.79 (10.48)                               | 0.077 | .000 |

All p-values are false discovery rate (FDR)-corrected p-values. Definite overreporting cutoff values are as follows: F-r ≥ 100; Fp-r ≥ 100; Fs ≥ 100; FBS-r ≥ 100; RBS ≥ 100; L-r ≥ 80; K-r: ≥ 70

RAVLT Rey Auditory Verbal Learning Test, BVMT-R Brief Visuospatial Memory Test-Revised, WAIS-IV PSI Wechsler Adult Intelligence Scale-Fourth Edition processing speed index

* p < .05

A Significant difference between 0 overreporting elevations 1 overreporting elevations

B Significant difference between 0 overreporting elevations and ≥ 2 overreporting elevations
Furthermore, the inclusion of two non-memory-based PVTs among the five criterion PVTs eliminates the potential confound of sole reliance on memory-based PVTs, which has been pervasive in the SVT/PVT literature. That said, one of the major methodological limitations was that only a single embedded validity indicator was included in the reference standard to avoid criterion contamination when analyzing commonly administered cognitive measures. This may inadvertently constrain the generalizability of the current results as most practicing neuropsychologists tend to include multiple embedded validity indicators throughout their evaluations to allow for continuous sampling of performance validity. As for other limitations, the use of a fairly heterogeneous neuropsychiatric sample such as this can fail to delineate potential confounds related to the adverse impact of more severe cognitive impairment on PVT performance, potentially obscuring its relationship with SVT symptom reporting. Furthermore, this sample contained a subset of actively compensation-seeking patients, which tends to be associated with higher rates of invalidity than patient populations without evidence of external incentives and likely influenced the frequency of failed PVTs and invalid symptom reporting.

Future research should further evaluate how the relationship between MMPI-2-RF symptom validity scales may vary depending on reference standards consisting of different combinations of freestanding and embedded PVTs. This study used MMPI-2-RF content-based validity scales as SVTs, whereas other SVTs, such as those from the PAI, may differentially relate to PVT and cognitive performance, further research may benefit from addressing how performance validity may relate to symptom reporting on the MMPI-2-RF, valid or otherwise.

Taken together, this study offers further support for utilizing both SVTs and PVTs in differentiating credible from noncredible self-report and cognitive performance, respectively, as part of a comprehensive evaluation for patients presenting with a variety of neurological and neuropsychiatric concerns. These findings provide additional evidence supporting the notion that PVTs and SVTs are dissociable constructs and offer unique, nonredundant information within neuropsychological evaluations. As such, both PVTs and SVTs serve as critical tools to help ascertain the accuracy of test results and symptom reporting, respectively, thereby providing independent information to increase the confidence in and accuracy of clinical decision-making.

Declarations

Conflict of Interest The authors declare no competing interests.

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/aca053

American Academy of Clinical Neuropsychology. (2007). American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. The Clinical Neuropsychologist, 21, 209–231. https://doi.org/10.1080/13855860601025932

Bailey, K. C., Soble, J. R., Bain, K. M., & Füllen, 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, 895–900. https://doi.org/10.1080/13854046.2016.1245787

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for Beck Depression Inventory (2nd ed.). Psychological Corporation.

Benedict, R. H. B. (1997). Brief Visuospatial Memory Test-Revised. Psychological Assessment Resources.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Ben-Porath, Y. S. (2012). Interpreting the MMPI-2-RF. University of Minnesota Press.

Ben-Porath, Y. S., & Tellegen, A. (2008). MMPI-2-RF: Manual for administration, scoring, and interpretation. University of Minnesota Press.

Bernstein, M. T., Resch, Z. J., Osview, G. P., & Soble, J. R. (2021). A systematic review and meta-analysis of the diagnostic accuracy of the Advanced Clinical Solutions Word Choice Test as a performance validity test. Neuropsychology Review, 31(2), 349–359. https://doi.org/10.1007/s11065-020-09468-y

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 examination. The Clinical Neuropsychologist, 23(4), 729–741. https://doi.org/10.1080/13854040904247803

Boone, K. B. (2013). Clinical practice of forensic neuropsychology: An evidenced-based approach. Guilford Press.

Boone, K. B., Lu, P., & Wen, J. (2005). Comparison of various RAVLT scores in the detection of noncredible memory performance. Archives of Clinical Neuropsychology, 20(3), 301–319. https://doi.org/10.1016/j.acn.2004.08.001

Boone, K. B., Lu, P., & Herzberg, D. (2002). The Dot Counting Test manual. Western Psychological Services.

Bracken, B., & Boatwright, B. (2005). CAT-C, Clinical Assessment of Attention Deficit-Child and CAT-A, Clinical Assessment of Attention Deficit-Adult professional manual. Lutz, FL: Psychological Assessment Resources.

Bush, S., Ruff, R., Troster, A., Barth, J., Koffler, S., Pilsner, N., et al. (2005). Symptom validity assessment: Practice issues and medical necessity – NAN Policy and Planning Committee. Archives of Clinical Neuropsychology, 20, 419–426. https://doi.org/10.1016/j.acn.2005.02.002

Copeland, C. T., Mahoney, J. J., Block, C. K., Linck, J. F., Pastorek, N. J., Miller, B. I., Romesser, J. M., & Sim, A. H. (2015). Relative

© Springer

70 Psychological Injury and Law (2023) 16:61–72
utility of performance and symptom validity tests. Archives of Clinical Neuropsychology, 31(1), 18–22. https://doi.org/10.1093/acnvn/ovx065

Fokas, K. F., & Brovko, J. M. (2020). Assessing symptom validity in psychological injury evaluations using the MMPI-2-RF and the PAI: An updated review. Psychological Injury and Law, 13(4), 370–382. https://doi.org/10.1080/21207-020-09393-8

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

Gervais, R. O., Wygant, D. B., Sellhom, M., & Ben-Porath, Y. S. (2011). Associations between symptom validity test failure and scores on the MMPI-2-RF validity and substantive scales. Journal of Personality Assessment, 93(5), 508–517. https://doi.org/10.1080/00223891.2011.594132

Golden, C. J. (1978). Green’s medical symptom validity test (MSVT) for the MMPI-2. Psychological Assessment, 1(4), 625–630. https://doi.org/10.1037/1073-1911.12098391

Ingram, P. B., Golden, B. L., & Armistead-Jehle, P. J. (2020). Evaluating the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) over-reporting scales in a military neuropsychology clinic. Journal of Clinical and Experimental Neuropsychology, 42(3), 263–273. https://doi.org/10.1080/13833395.2019.1708271

Ingram, P. B., Tarescavage, A. M., Ben-Porath, Y. S., & Oehlert, M. E. (2019). Descriptive profiles of the MMPI-2-Restructured Form (MMPI-2-RF) across a national sample of four veteran affairs treatment settings. Journal of Psychopathology and Behavioral Assessment, 41(2), 329–340. https://doi.org/10.1007/s10862-019-09727-0

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

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. (2021). 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, 1–19. https://doi.org/10.1080/13854046.2021.1900398

Khan, H., Rauch, A. A., Obolsky, M. A., Skymba, H., Barwegen, K. C., Wisinger, A. M., Ovsiew, G. P., Jennette, K. J., Soble, J. R., & Resch, Z. J. (2022). A comparison of embedded validity indicators from the Stroop Color and Word Test among adult referrals for clinical evaluation of suspected or confirmed attention-deficit/hyperactivity disorder. Psychological Assessment. https://doi.org/10.1037/pas0001137

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. The Clinical Neuropsychologist, 22(4), 666–679. https://doi.org/10.1080/13854040401494987

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. Journal of the International Neuropsychological Society, 18(4), 625–630. https://doi.org/10.1017/s1355617712000240

Leib, S. I., Schieszer-Ockrassa, C., White, D. J., Gallagher, V. T., Carter, D. A., Basurto, K. S., Ovsiew, G. P., Resch, Z. J., Jennette, K. J., & Soble, J. R. (2021). Concordance between the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) and Clinical Assessment of Attention Deficit-Adult (CAT-A) over-reporting validity scales for detecting invalid ADHD symptom reporting. Applied Neuropsychology: Adult, 1–8. https://doi.org/10.1080/23729905.2021.1894150

Martin, P. K., Schroeder, R. W., Olsen, D. H., 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. https://doi.org/10.1080/13854046.2019.1637027

Morgan, L. C. (1991). The Personality Assessment Inventory. Odessa, FL: Psychological Assessment Inventory.

Neale, A. C., Ovsiew, G. P., Resch, Z. J., & Soble, J. R. (2022). Feigning or forgetfulness: The effect of memory impairment severity on Word Choice Test performance. The Clinical Neuropsychologist, 36(3), 584–599. https://doi.org/10.1080/13854046.2020.1799076

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. https://doi.org/10.1037/neut0000722

Pearson, N. (2009). Advanced clinical solutions for WAIS-IV and WMS-IV: Administration and scoring manual. The Psychological Corporation.

Pliskin, J. I., De-Dios-Stern, S., Resch, Z. J., Saladino, K. F., Ovsiew, G. P., Carter, D. A., & Soble, J. R. (2021). Comparing the psychometric properties of 8 embedded performance validity tests in the Rey Auditory Verbal Learning Test, Wechsler Memory Scale, Logical Memory, and Brief Visuospatial Memory Test-Revised recognition trials for detecting invalid neuropsychological test performance. Assessment, 28(8), 1871–1881. https://doi.org/10.1080/13854046.2021.1947218

Resch, Z. J., Rhoads, T., Ovsiew, G. P., & Soble, J. R. (2022b). A known-groups validation of the Medical Symptom Validity Test and analysis of the Genuine Memory Impairment Profile. Assessment, 29(3), 455–466. https://doi.org/10.1177/1073191120983919

Rhoads, T., Neale, A. C., Resch, Z. J., Cohen, C. D., Kezer, R. D., Cerny, B. M., Jennette, K. J., Ovsiew, G. P., & Soble, J. R. (2021). Psychometric implications of failure on one performance validity test: A cross-validation study to inform criterion group definition. Journal of Clinical and Experimental Neuropsychology, 43(5), 437–448. https://doi.org/10.1080/13833095.2021.1945540

Schmidt, M. (1996). Rey Auditory Verbal Learning Test: A handbook. Western Psychological Services.

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. Assessment, 19(1), 21–30. https://doi.org/10.1177/1073191111428764

Shura, R. D., Yoash-Gantz, R. E., Pickett, T. C., McDonald, S. A., & Tupper, L. A. (2021). Relations among performance and symptom validity, mild traumatic brain injury, and posttraumatic stress disorder symptom burden in postdeployment veterans. Psychological Injury and Law, 14, 257–268. https://doi.org/10.1007/s12207-021-09415-z

Soble, J. R., Critchfield, E. A., & O’Rourke, J. J. F. (2017). Neuropsychological evaluation in traumatic brain injury. Physical Medicine and Rehabilitation Clinics of North America, 28(2), 339–350. https://doi.org/10.1016/j.pmar.2016.12.009

Soble, J. R., Resch, Z. J., Schulze, E. T., Paxton, J. L., Cation, B., Friedhoff, C., Costin, C., Fink, J. W., Lee, R. C., & Pliskin, N. 71
H. (2019). Examination of the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) validity and substantive scales in patients with electrical injury. *The Clinical Neuropsychologist, 33*(8), 1501–1515. https://doi.org/10.1080/13854046.2019.1616114

Soble, J. R., Sharp, D. W., Carter, D. A., Jennette, K. J., Resch, Z. J., Ovsiew, G. P., & Critchfield, E. A. (2021a). Cross-validation of a forced-choice validity indicator to enhance the clinical utility of the Rey Auditory Verbal Learning Test. *Psychological Assessment, 33*(6), 568–573. https://doi.org/10.1037/pas00001018

Soble, J. R., Silva, M. A., Vanderploeg, R. D., Curtiss, G., Belanger, H. G., Donnell, A. J., & Scott, S. G. (2014). Normative Data for the Neurobehavioral Symptom Inventory (NSI) and post-concussion symptom profiles among TBI, PTSD, and nonclinical samples. *The Clinical Neuropsychologist, 28*(4), 614–632. https://doi.org/10.1080/13854046.2014.894576

Soble, J. R., Webber, T. A., & Bailey, K. C. (2021b). An overview of common stand-alone and embedded PVTs for the practicing clinician: Cutoffs, classification accuracy, and administration times. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 126–149). Guilford.

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., & Participants, C. (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*, 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

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

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. *The Journal of Head Trauma Rehabilitation, 29*(1), 1–10. https://doi.org/10.1097/HTR.0b013e318281966e

Wechsler, D. (2008). *WAIS-IV: Administration and scoring manual*. The Psychological Corporation.

White, D. J., Ovsiew, G. P., Resch, Z. J., & Soble, J. R. (2020). Cross-validation of non-memory-based embedded performance validity tests for detecting invalid performance among patients with and without neurocognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 42*(5), 459–472. https://doi.org/10.1080/13803395.2020.1758634

White, D. J., Ovsiew, G. P., Rhoads, T., Resch, Z. J., Lee, M., Oh, A. J., & Soble, J. R. (2022). The divergent roles of symptom and performance validity in the assessment of ADHD. *Journal of Attention Disorders, 26*(1), 101–108. https://doi.org/10.1177/1087054720964575

Whitney, K. A., Davis, J. J., Shepard, P. H., & Herman, S. M. (2008). Utility of the response bias scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Archives of Clinical Neuropsychology, 23*(7–8), 777–786. https://doi.org/10.1016/j.acn.2008.09.001

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.