Applied Neuropsychology: Adult

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hapn21

Detecting malingered neurocognitive dysfunction: Comparative analysis of freestanding and embedded performance validity tests

Sonia Sistiaga, Sarah Gilis, Perrine Wilmotte, Audrey Vicenzutto & Isabelle Simoes Loureiro

To cite this article: Sonia Sistiaga, Sarah Gilis, Perrine Wilmotte, Audrey Vicenzutto & Isabelle Simoes Loureiro (19 Sep 2024): Detecting malingered neurocognitive dysfunction: Comparative analysis of freestanding and embedded performance validity tests, Applied Neuropsychology: Adult, DOI: 10.1080/23279095.2024.2404195

To link to this article: https://doi.org/10.1080/23279095.2024.2404195

APPLIED NEUROPSYCHOLOGY: ADULT
https://doi.org/10.1080/23279095.2024.2404195

Detecting malingered neurocognitive dysfunction: Comparative analysis of freestanding and embedded performance validity tests

Sonia Sistiaga[a], Sarah Gilis[a], Perrine Wilmotte[a], Audrey Vicenzutto[b] and Isabelle Simoes Loureiro[a]

[a]Cognitive Psychology and Neuropsychology Department, University of Mons, Mons, Belgium; [b]Forensic Psychology Department, University of Mons, Mons, Belgium

ABSTRACT

Detecting malingered neurocognitive dysfunction is a major issue in forensic context, particularly in legal proceeding/insurance assessment after a traumatic brain injury (TBI), condition frequently associated with persistent cognitive impairments that may potentially be related to malingering. Consequently, research has devoted considerable efforts on developing tools to verify symptoms authenticity. This study compared two freestanding performance validity tests (PVTs) (Amsterdam Short-Term Memory Test—ASTM; Word Completion Memory Test—WCMT) and five embedded PVTs (Rey Complex Figure Test—RCFT—Copy and Recall trials; Reliable Digit Span—RDS; Rey Auditory Verbal Learning Test—RAVLT—Recognition and Total Learning trials) in a sample of 120 participants, including 15 patients with TBI (M_{age} = 44.40), 52 experimental simulators (M_{age} = 29.52) and 53 control (M_{age} = 29.77). Group performance was analyzed to assess tests' discriminatory power, and Receiver Operating Characteristic (ROC) curves were used to examine tools' sensitivity and specificity. Results indicated that experimental simulators performance on the ASTM, WCMT, and RAVLT differed significantly from TBI patients and controls. The RDS and RCFT did not discriminate experimental simulators from TBI group. ROC curves analysis reveals that the most accurate tests in this battery for detecting malingering were the ASTM and the RAVLT-Total Learning score. This study offers implications for identifying malingering in medico-legal settings, underscoring the importance of incorporating PVTs into clinical practice.

KEYWORDS

Forensic expertise; malingering; neuropsychological assessment; performance validity; traumatic brain injury

Introduction

In clinical assessment, patient behavior is essential for accurate diagnosis and effective intervention recommendations. Some patients, often claimants, might exaggerate or simulate their disorders to gain secondary benefits, such as financial compensation or avoiding criminal prosecution (Barthélémy et al., 2014). These patients may inflate the frequency, intensity, severity, and/or duration of their symptoms to influence the results of their examination (Iverson et al., 2007), mention false symptoms, or show unusual symptoms linked to a psychodynamic bias. Therefore, the justice system seeks scientifically grounded evidence to ensure fair and informed decisions, exploring various fields and methodologies, including neuropsychology (Serafim et al., 2015). Neuropsychological assessment assumes that the individual provides truthful responses and adequate test effort (Bush et al., 2014; Millis, 2009). Insufficient test effort can occur for several reasons, whether deliberate (e.g., malingering) or unintentional (e.g., lack of motivation or commitment), compromising the validity of tests results (Schroeder et al., 2016). Undetected invalid assessments can lead to adverse consequences, such as misdiagnosis, inappropriate intervention strategies, incorrect data in treatment efficacy studies, and unfair allocation of resources and financial compensation (Roor et al., 2022). These challenges notably extend to conditions like traumatic brain injury (TBI) and their nuanced neuropsychological assessments. The TBI population is diverse and complex, with varying causes and severity levels of injury, leading to a range of cognitive and emotional consequences (Menon et al., 2010). Forensic psychological assessment following TBI plays a significant role in determining trauma-related disability compensation. Estimates of malingering among head injury litigants vary widely, reportedly ranging from 1 to 50% (Larrabee et al., 2009; Mittenberg et al., 2002; Reynolds, 1998). Additionally, the prevalence of cognitive underperformance is estimated at 15–30% in TBI patients (Donders & Boonstra, 2007; Moore & Donders, 2004; Stulemeijer et al., 2007). Performance validity tests (PVTs) have thus been developed to objectively evaluate the truthfulness of clinical presentations (Barthélémy et al., 2014). There are two types of PVTs: freestanding PVTs, which are specifically designed to assess performance invalidity, and embedded PVTs, which are classic neuropsychological tests with established cutoff scores to detect malingering. Many experts recommend using both freestanding and embedded tests (Bush et al., 2005; Heilbronner

CONTACT Sonia Sistiaga soniasistiaga@gmail.com Cognitive Psychology and Neuropsychology Department, University of Mons, Mons, Belgium.
Supplemental data for this article is available online at https://doi.org/10.1080/23279095.2024.2404195

et al., 2009), a guideline also supported by the Slick criteria (Slick et al., 1999).

This study aims to compare the performance of several freestanding and embedded PVTs. We utilized two freestanding PVTs. The Amsterdam Short-Term Memory Test (ASTM; Schagen et al., 1997), widely recognized and used in malingering research, has been extensively validated across 17 normative and clinical patient groups (Schmand et al., 2005), demonstrating its sensitivity and validity in assessing underperformance (Schagen et al., 1997). Validity studies indicate that a cutoff score of 86 is associated with a specificity of 83% and a sensitivity of 92% (Schmand et al., 2005). Based on these studies, its internal consistency has been shown to be excellent. The Word Completion Memory Test (WCMT; Hilsabeck et al., 2001), although receiving less attention, is presented as a valid and effective tool for detecting malingering. The WCMT was designed to address the difficulties other tests faced in detecting simulators with extensive knowledge of neuropsychological disorders. In its validation study, the WCMT achieved an overall specificity of 97.2% and exhibited strong psychometric properties, making it a promising tool for identifying simulated memory disorders (Hilsabeck & Gouvier, 2005). Additionally, we included five widely recognized embedded PVTs derived from three instruments. First, from the Rey Complex Figure Test (RCFT; Meyers & Meyers, 1995), we have focused our analysis on the Copy and Immediate-Recall scores. Numerous groups have suggested that raw scores from the RCFT-Copy and Immediate-Recall trials can be used as PVTs. Reedy et al. (2013) classified participants into credible and non-credible performance groups based on their PVT performance and litigation status. They observed good discrimination with ≥91% specificity and ≥41% sensitivity based on cutoff scores from the Copy and Immediate-Recall trials. Second, we have used the Reliable Digit Span (RDS; Greiffenstein et al., 1994), which is a well-validated embedded validity indice (EVI), with a cutoff score of 7 typically indicating suboptimal effort (Greiffenstein et al., 1994; Meyers & Volbrecht, 1998; Schroeder et al., 2012). However, a systematic review by Schroeder et al. (2012) revealed that the 7 cutoff score resulted in inadequate specificity across clinical groups, whereas a cutoff of 6 achieved specificity rates above 90%. Recent studies have corroborated these findings, suggesting that a cutoff of 6 generally performs better (Maiman et al., 2018; Zenisek et al., 2016). Finally, from the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1958), we used the Total Learning and Recognition scores. Suhr and Barrash (2007) identified 18 different variables that have been examined in studies using simulation designs or clinical groups, with the majority utilizing simulation designs. Despite the extensive volume of studies, there is a lack of consensus among reviewers regarding the most efficient variables (Suhr & Barrash, 2007). The exception appears to be the Recognition trial, which has demonstrated utility across multiple studies, with different raw cutoff scores: 6 (Binder et al., 2003) or 10 (Meyers et al., 2001). In our study, we examined the Total Learning and Recognition scores.

In sum, the main purpose of this study is to compare two freestanding PVTs and five EVIs among a French-speaking sample of experimental malingering simulators, TBI patients, and controls. As research in the field of malingering is limited in French, studies involving multiple PVTs are particularly valuable for assessing consistency across measures and enhancing measurement reliability (Axelrod & Schutte, 2011).

Method

Participants and procedure

A total of 140 participants were recruited for this study through advertisements on social networks. Each participant signed an informed consent form in accordance with the Declaration of Helsinki and the General Data Protection Regulation (GDPR) (European Parliament, 2016) on protecting personal data. The current study received a favorable opinion by the Ethics Committee of the first author's faculty. Participants were selected based on the following exclusion criteria: non-French speaker, history of neurological problems (except for the TBI group), current major medical condition, psychiatric condition, past or current substance/alcohol dependence, neuropsychological testing within the last three months. Eleven participants were excluded due to elevated levels of anxiety and depression, three for medical history, and six for neurodevelopmental disorders. The final sample of 120 participants (18–60 years, M_{age} = 31.49 years, SD = 13.79, N = 82 women) included 15 patients (M_{age} = 44.40, SD = 11.67; $M_{years\ of\ education}$ = 13.13, SD = 8.23; N = 8 women) with TBI, 52 subjects (M_{age} = 29.52, SD = 13.56; $M_{years\ of\ education}$ = 14.80, SD = 8.11; N = 35 women) instructed to credibly simulate to suffer from cognitive impairment following TBI to obtain money (experimental simulators), and 53 control participants (M_{age} = 29.77, SD = 12.76; $M_{years\ of\ education}$ = 13.40, SD = 8.23; N = 39 women). Patients from TBI group provided medical evidence of the injury, which also indicated the level of trauma severity. The TBI group comprised an equal distribution of levels of severity, with five mild ($M_{number\ of\ years\ since\ the\ injury}$ = 4.20; SD = 5.49), five moderate ($M_{number\ of\ years\ since\ the\ injury}$ = 6.60; SD = 3.05), and five severe cases ($M_{number\ of\ years\ since\ the\ injury}$ = 8.00; SD = 5.00).

Testing was completed in a single session. TBI patients were instructed to put forth their full effort on all measures administered. We followed the procedure commonly adopted by studies on the topic involving experimental malingering simulators (Jelicic et al., 2011; Zasler & Bigler, 2017): participants received a letter (see Supplementary Method S1) containing information on the symptoms associated with TBI and instructions on how to simulate the symptoms as credibly as possible, without exaggerating them. Participants in the control group received a letter encouraging them to perform the tests to the best of their ability. The study was conducted in a single-blind setting: the investigator did not know whether the participants were going to simulate or not.

Measures

Freestanding performance validity tests

Amsterdam Short-Term Memory Test (ASTM; Schagen et al., 1997) is a forced-choice test of verbal memory. The ASMT consists of 30 evaluation items. The subject is asked to read

aloud five words from the same semantic category and memorize them. Next, a short distracting task involving simple addition or subtraction calculations (e.g. 50–25) is presented, followed by a recognition phase (five words: three target words and two low-prototype distractors). When scoring, each recognized word adds one point to the total score, with a maximum score of 90. We used the French validation of the computerized version in our study (Meulemans et al., 2003).

Word Completion Memory Test (WCMT; Hilsabeck et al., 2001) is a measure of implicit memory priming. The WCMT consists of two subtests. In the Inclusion subtest, the participant has to memorize a list of 30 words. The participant is instructed to read each word aloud, spell it out, and write it down, indicating whether the word is "pleasant" or "unpleasant." The participant is then asked to complete 30 word starts, with the aim of returning the learned list. The Exclusion task is similar, except that the participant must avoid completing the list of 30 word starts with those from the previous list. When scoring, each word completed in accordance with the instructions is worth one point. The score is between 0 and 30 for the inclusion task (Score I) and between 0 and 30 for the exclusion task (Score E). The R score, supposed to measure participants' degree of conscious control over verbal memorization (Hilsabeck et al., 2001), is obtained by subtracting the I score from the E score (I−E=R). The score ranges from −30 to 30. The more negative the score, the greater the suspicion of malingering in the assessment.

Embedded performance validity tests

Rey Complex Figure Test (RCFT; Meyers & Meyers, 1995) is a neuropsychological test assessing working memory, visual long-term memory, attention, and praxis. The test presents a geometric figure made up of 18 elements organized into three parts: a global shape, external elements, and elements internal to the overall shape. As embedded indices, we focused our analysis on the Copy and Immediate Recall (i.e., three min-delayed) scores.

Reliable Digit Span (RDS; Greiffenstein et al., 1994) obtained from the Digit Span subtest (WAIS-IV; French version, Wechsler, 2011), is among the most commonly used embedded PVTs in neuropsychology (Boone, 2007). RDS is calculated by adding the maximum number of digits forward with the maximum number of backward when both trials were passed (Greiffenstein et al., 1994).

Rey Auditory Verbal Learning Test (RAVLT; Rey, 1958) is designed to assess the encoding, storage, and retrieval of auditory-verbal information in long-term memory. Participants are presented with a 15-word list (list A) five times, each followed by an attempted recall. This is followed by the presentation of a second 15-word interference list (list B), and subsequent recall of list A. Recognition is tested through the Recognition score. We focused our analysis on the Total Learning and Recognition scores.

Data analysis

Descriptive analysis

Analyses were performed using IBM SPSS Statistics 27.0 (Armonk, NY, USA) statistical software. The means and standard deviations of the total scores for the tools described above were computed. The absence of outliers allowed us to use the raw scores from our PVTs and not be forced to use *Z*-scores or percentiles (Rogers & Bender, 2020).

Group performance comparison

Mixed analyses of variance (ANOVAs) were computed with "Group" as a between-subject factor. Post hoc Bonferroni-corrected *t*-tests and supplementary analyses—multivariate ANOVAs—were used to accordingly disentangle significant main effect (significance level $p < .05$). The influence of age and gender on test performance was controlled and demonstrated no significant effect.

The anonymized dataset is available on the Open Science Framework: https://osf.io/u52jp/?view_only=f6e65fe9ea7d415f9a202383183f3c7b.

Measurement of sensitivity and reliability

Receiver Operating Characteristics (ROC) curves were used to analyze the power of our PVTs to detect malingered neurocognitive dysfunction. Area under the curve (AUC) values above .90 indicate "outstanding discrimination," between .80 and .89 indicate "excellent discrimination (Hosmer & Lemeshow, 2000), and below suggest relatively low-test precision. Sensitivity refers to the test's ability to correctly predict that a simulator has indeed malingered (true positives), while specificity is the test's ability to correctly predict that a non-simulator has not, in fact, malingered (true negatives). Using ROC curves, we reported the sensitivity and specificity value for the cutoff score established in the literature, based on our sample.

Complementary analysis: number of failed performance validity tests

Following the Slick et al. (1999) criteria, which emphasize the importance of multiple PVTs failures to assess the likelihood of malingering, we classified each participant's performance on every PVT as either failed or passed, based on the literature-established cutoff scores. We then calculated the total number of failed PVTs per participant.

Results

Group performance comparison

Freestanding PVTs

The comparative analysis of performance between our groups on freestanding PVTs is presented in Table 1.

Results indicate that the ASTM distinguishes experimental simulators significantly from controls and TBI patients. The WCMT Inclusion and R scores enable to differentiate between experimental simulators, control, and TBI groups, as well as between controls and TBI patients.

Embedded PVTs

The comparative analysis of performance between the groups on embedded PVTs is presented in Table 2.

Table 1. Group performance comparison for freestanding performance validity tests.

	Anova			Bonferroni's *post-hoc*				
Freestanding PVTs	*F*(2,77)	*p*	Partial η^2	Group	Mean (*SD*)	ES *vs.* C	ES *vs.* TBI	C *vs.* TBI
ASTM total	72.496	<.001	.553	ES	75.33 (8.48)	<.001	<.001	1.000
				TBI	87.93 (1.94)			
				C	88.28 (2.09)			
WCMT I	41.737	<.001	.416	ES	13.85 (5.19)	<.001	<.001	.003
				TBI	24.27 (3.90)			
				C	19.85 (3.77)			
WCMT R	20.203	<.001	.257	ES	1.00 (10.86)	.002	<.001	<.001
				TBI	22.60 (5.89)			
				C	9.23 (14.01)			

PVTs: performance validity tests; ES: experimental simulators ($N = 52$); TBI: traumatic brain injury ($N = 15$); C: controls ($N = 53$); ASTM: Amsterdam Short-Term Memory Test; WCMT: World Completion Memory Test; I: inclusion task; R: R score (Inclusion − Exclusion).

Table 2. Group performance comparison for embedded performance validity tests.

	ANOVAs			Bonferroni's *post-hoc*				
Embedded PVTs	*F*(2,77)	*p*	Partial η^2	Group	Mean (*SD*)	ES *vs.* C	ES *vs.* TBI	C *vs.* TBI
RCFT-Copy	4.610	.012	.074	ES	33.88 (3.00)	.009	.694	1.000
				TBI	34.67 (2.02)			
				C	35.21 (1.15)			
RCFT-Recall	14.655	<.001	.203	ES	20.22 (5.09)	<.001	1.000	.009
				TBI	20.93 (4.45)			
				C	25.25 (4.77)			
RDS	27.551	<.001	.324	ES	6.46 (1.85)	<.001	1.000	<.001
				TBI	6.20 (1.08)			
				C	9.06 (2.18)			
RAVLT-Total Learning	38.608	<.001	.398	ES	42.60 (12.21)	<.001	<.001	1.000
				TBI	57.80 (7.59)			
				C	58.64 (7.39)			
RAVLT-Recognition	4.661	.011	.074	ES	10.40 (4.68)	1.000	.018	.012
				TBI	15.00 (0.00)			
				C	10.15 (7.05)			

PVTs: performance validity tests; ES: experimental simulators ($N = 52$); TBI: traumatic brain injury ($N = 15$); C: controls ($N = 53$); RCFT: Rey Complex Figure Test; RDS: Reliable Digit Span; RAVLT: Rey Auditory Verbal Learning Test.

Table 3. ROC curve analysis of freestanding performance validity tests in malingered neurocognitive dysfunction detection.

Freestanding PVTs	Suggested cutoff score	Associated specificity	Associated sensitivity	Associated specificity in our sample	Associated sensitivity in our sample	AUC (*p*)
ASTM Total	≤86 (Schmand et al., 2005)	92	83	86.8	86.5	.951 (<.001)
WCMT I	≤15 (Hilsabeck et al., 2001)	100	86	98.1	14.7	.852 (<.001)
WCMT R	≤9 (Hilsabeck et al., 2001)	100	93	75	69.1	.749 (<.001)

PVTs: performance validity tests; AUC: area under the curve; ASTM: Amsterdam Short-Term Memory Test; WCMT: World Completion Memory Test; I: inclusion task; R: R score (Inclusion − Exclusion).

Results indicate that the RCFT-Copy score only differentiates experimental simulators from controls. The RCFT-Recall score distinguishes experimental simulators from controls but also controls from TBI patients. The RDS differentiates experimental simulators from controls but not from TBI patients. The RAVLT-Total Learning score distinguishes between experimental simulators, controls, and TBI patients, but not yet between controls and TBI patients. The Recognition score differentiates experimental simulators from TBI patients but not from controls.

ROC curves analysis

Freestanding PVTs

ROC curve analysis and associated indices (AUC, specificity, sensitivity) of freestanding PVTs are presented in Table 3.

The AUC values fall in the range of "outstanding discrimination" (≥.90; Hosmer & Lemeshow, 2000) for the ASTM, "excellent discrimination" (.80–.89) for the WCMT I score, but were lower for the WCMT R score. In our sample, the specificity and sensitivity values associated with the literature-established cutoff scores were lower than those reported in the original studies. Only the WCMT I score reached the 90% specificity threshold.

Embedded PVTs

ROC curve analysis and associated indices (AUC, specificity, sensitivity) of embedded PVTs are presented in Table 4.

The AUC values fall in the range of "excellent discrimination" (.80–.89) for the RAVLT-Total Learning score but were lower for the RDS, the RCFT-Copy and Recall scores, and the RAVLT-Recognition score. In our sample, the specificity and sensitivity values associated with the literature-established cutoff scores were higher for the RCFT-Copy and Recall scores and for the RAVLT-Total Learning score, but lower for the RDS and RAVLT-Recognition

Table 4. ROC curve analysis of embedded performance validity tests in malingered neurocognitive dysfunction detection.

Embedded PVTs	Suggested cutoff score	Associated specificity	Associated sensitivity	Associated specificity in our sample	Associated sensitivity in our sample	AUC (*p*)
RCFT copy	≤26 (Reedy et al., 2013)	90	52	100	58	.609 (.040)
RCFT recall	≤10 (Reedy et al., 2013)	88.3	45	100	19	.721 (<.001)
RDS	≤7 (Greiffenstein et al., 1994)	73	70	72.1	60	.764 (<.001)
RAVLT total learning	≤30 (Boone et al., 2005)	96.0	42.6	100	18	.865 (<.001)
RAVLT recognition	≤10 (Boone et al., 2005)	88	77	75	36	668 (.002)
	≤6 (Binder et al., 2003)	95	38	75	18	

PVTs: performance validity tests; AUC: area under the curve; RCFT: Rey Complex Figure Test; RDS: Reliable Digit Span; RAVLT: Rey Auditory Verbal Learning Test.

Table 5. Distribution of participants failing performance validity tests.

Group	Number of participants failing 2 or more PVTs	Number of participants failing 3 or more PVTs
Experimental simulators (*N* = 52)	48 (92.31%)	42 (80.77%)
Controls (*N* = 53)	19 (35.85%)	4 (7.55%)
Traumatic brain injury (*N* = 15)	7 (46.67%)	0 (0%)

scores compared to the original studies. The RCFT-Copy, Recall, and RAVLT-Total Learning scores met the 90% specificity threshold in our sample.

Complementary analysis: number of failed performance validity tests

The distribution of the number of failed PVTs in each group is presented in Table 5.

Results indicate that among the experimental simulators, 48 participants failed two or more PVTs, and 42 participants failed three or more PVTS. In the control group, 19 participants failed two or more PVTs, while only four participants failed three or more PVTs. In the TBI group, seven participants failed two more PVTs, but none failed three or more PVTs.

Discussion

The main aim of this study was to compare the performance of different tools in the detection of malingered neurocognitive dysfunction. For this purpose, we used two freestanding PVTs (the ASTM and the WCMT) and five embedded PVTs (the RCFT-Copy and Recall scores; the RDS; the RAVLT-Total Learning and Recognition scores). The performances of 52 experimental malingering simulators, 53 controls, and 15 patients with TBI were compared.

Regarding freestanding PVTs, the ASTM correctly distinguished experimental simulators from control and TBI patients. ROC curves report high values for AUC and specificity, attesting to good accuracy in malingering detection. The WCMT I and R scores correctly discriminated between all groups. AUC values were also high for the WCMT I score, suggesting good accuracy in malingering detection. Concerning EVIs, the majority of the tests we included in our study correctly differentiated experimental simulators from controls, but not from TBI patients. The RAVLT-Total Learning score was the only one to differentiate experimental simulators from both control and TBI patients. The psychometric qualities associated with ROC curve analysis were more heterogeneous. Again, the RAVLT-Total Learning score was the only one to show an AUC value above .80, attesting to good precision in the detection of malingering, and the specificity value reached 100% in our sample. Indeed, to avoid false identification of the malingering, PVTs have been designed to produce high specificity, at the expense of relatively low sensitivity (Bush et al., 2005; Heilbronner et al., 2009; Larrabee, 2008). Favoring the reduction of false positives seems essential, compared with false negatives, to avoid unfairly depriving an individual of resources and compensation to which he or she is entitled (Boone, 2007; Iverson et al., 2007; Larrabee, 2008). Researchers strive to keep false-positive rates below 10% (Boone et al., 2013; Larrabee, 2014). AUC values were lower for the RCFT-Copy and Recall scores, but the RCFT-Copy score also demonstrated a specificity value of 100% in our sample. In contrast, the RDS showed a low AUC value, and its specificity did not reach the 90% threshold in our sample.

Similar performances between TBI patients and controls were mostly observed for the RCFT-Copy and the RAVLT-Total Learning scores. However, it would not have been surprising to find significant differences between these groups. While many TBI patients experience full neurologic recovery, some may develop prolonged neurocognitive and behavioral changes (Daneshvar et al., 2011; Wortzel & Arciniegas, 2012). Cognitive consequences vary depending on the injury severity and brain lesion's location. Therefore, it may be useful to examine each patient's test failures to see if they correspond to the specific areas of their brain injury.

Research using the Slick criteria (1999) showed the value of requiring multiple failures on PVTs to determine probabilities of malingering. Several thresholds have been proposed for detecting malingering, with some authors suggesting that failing two or more PVTs provides sufficient specificity (Larrabee, 2003), while others advocate for a higher threshold of three or more PVT failures (Victor et al., 2009). Therefore, we calculated the total number of failed PVTs within our groups. The results suggest that in our sample, the threshold of three or more failures was more effective in distinguishing between experimental simulators and both controls and TBI patients, producing fewer false positives and thereby increasing specificity. However, PVT failures alert the examiner to the possibility of invalid data but do not reveal the underlying reason for the non-credible presentation (Boone et al., 2013). Indeed, some researchers have found no relationship between external incentives and PVT scores (Erdodi et al., 2018). Linking a PVT failure to

malicious intent lacks an epistemological basis and might not be ethical, especially considering the potentially severe adverse effects on the assessed individuals.

This study obviously has a few limitations. Our sample was predominantly female and included a relatively small sample of TBI patients. As such, the results cannot be generalized to different clinical populations (Erdodi & Rai, 2017; Glassmire et al., 2019). Similar studies with larger sample sizes, including patients with varying lesion loads and locations, and patients from different diagnostic categories are needed to gain a more general overview of the use of these tools in various clinical settings. Furthermore, the simulators were experimental simulators, i.e., healthy participants instructed to feign TBI-related cognitive impairment. We only had control over the instructions given to the experimental simulators and cannot attest the extent to which these instructions were followed. In addition, a significant portion of the control group (35.85%) failed at least two PVTs in our sample, raising further concerns about the effort and engagement of control participants in the testing session. In their study, Abeare et al. (2021) demonstrated that control participants failed the PVTs, while the experimental simulators produced intact cognitive profiles. Therefore, the validity of the group criteria cannot be asserted in our study.

Despite these limitations, it appears that freestanding PVTs are relatively robust to symptom coaching. However, it is important to consider that certain tasks may be selectively failed by patients not due to malingering, but because of neurologically based deficits. Therefore, while the combined use of freestanding and embedded PVTs should be considered when assessing cognitive dysfunction in forensic cases, the results must be carefully interpreted.

Acknowledgments

We thank Nathalie Ehrlé for her careful review of the manuscript. We thank Gabrielle Broze and Sabrina Ouirache, who contributed to the data collection. We thank all participants who took part in this study.

Author contributions

Sonia Sistiaga: methodology, formal analysis, writing–original draft. Sarah Gilis: methodology, formal analysis. Perrine Wilmotte: methodology, investigation. Audrey Vicenzutto and Isabelle Simoes Loureiro: conceptualization, methodology, writing–review and editing, supervision.

Disclosure statement

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

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Abeare, C. A., Hurtubise, J. L., Cutler, L., Sirianni, C., Brantuo, M., Makhzoum, N., & Erdodi, L. A. (2021). Introducing a forced choice Recognition trial to the Hopkins Verbal Learning Test – Revised. *The Clinical Neuropsychologist*, *35*(8), 1442–1470. https://doi.org/10.1080/13854046.2020.1779348

Axelrod, B. N., & Schutte, C. (2011). Concurrent validity of three forced-choice measures of symptom validity. *Applied Neuropsychology*, *18*(1), 27–33. https://doi.org/10.1080/09084282.2010.523369

Barthélémy, R., Lenne, B., Hautecœur, P., & Antoine, P. (2014). Cognitive disorders malingering: conceptual and methodological aspects. *Revue de Neuropsychologie*, *6*(2), 110–116. https://doi.org/10.1684/nrp.2014.0300

Binder, L. M., Kelly, M. P., Villanueva, M. R., & Winslow, M. M. (2003). Motivation and neuropsychological test performance following mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *25*(3), 420–430. https://doi.org/10.1076/jcen.25.3.420.13806

Boone, D., Mallett, S., Zhu, S., Yao, G. L., Bell, N., Ghanouni, A., von Wagner, C., Taylor, S. A., Altman, D. G., Lilford, R., & Halligan, S. (2013). Patients' & healthcare professionals' values regarding true- & false-positive diagnosis when colorectal cancer screening by CT colonography: Discrete choice experiment. *PLOS One*, *8*(12), e80767. https://doi.org/10.1371/journal.pone.0080767

Boone, K. B. (2007). *Assessment of feigned cognitive impairment. A neuropsychological perspective*. Guilford Publications.

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

Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the association for scientific advancement in psychological injury and law. *Psychological Injury and Law*, *7*(3), 197–205. https://doi.org/10.1007/s12207-014-9198-7

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

Daneshvar, D. H., Nowinski, C. J., McKee, A. C., & Cantu, R. C. (2011). The epidemiology of sport-related concussion. *Clinics in Sports Medicine*, *30*(1), 1–17, vii. https://doi.org/10.1016/j.csm.2010.08.006

Donders, J., & Boonstra, T. (2007). Correlates of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury*, *21*(3), 319–326. https://doi.org/10.1080/02699050701253129

Erdodi, L. A., & Rai, J. K. (2017). A single error is one too many: Examining alternative cutoffs on trial 2 on the TOMM. *Brain Injury*, *31*(10), 1362–1368. https://doi.org/10.1080/02699052.2017.1332386

Erdodi, L. A., Hurtubise, J. L., Charron, C., Dunn, A., Enache, A., McDermott, A., & Hirst, R. B. (2018). The D-KEFS trails as performance validity tests. *Psychological Assessment*, *30*(8), 1082–1095. https://doi.org/10.1037/pas0000561

European Parliament (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, *L 119*, 1–88. http://data.europa.eu/eli/reg/2016/679/oj

Glassmire, D. M., Wood, M. E., Ta, M. T., Kinney, D. I., & Nitch, S. R. (2019). Examining false-positive rates of Wechsler Adult Intelligence Scale (WAIS-IV) processing speed based embedded validity indicators among individuals with schizophrenia spectrum disorders. *Psychological Assessment*, *31*(1), 120–125. https://doi.org/10.1037/pas0000650

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*(3), 218–224. https://doi.org/10.1037/1040-3590.6.3.218

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

Hilsabeck, R. C., LeCompte, D. C., Marks, A. R., & Grafman, J. (2001). The Word Completion Memory Test (WCMT) a new test to detect

malingered memory deficits. *Archives of Clinical Neuropsychology*, *16*(7), 669–678. https://doi.org/10.1016/s0887-6177(00)00077-9

Hilsabeck, R., & Gouvier, W. (2005). Detecting simulated memory impairment: Further validation of the Word Completion Memory Test (WCMT). *Archives of Clinical Neuropsychology*, *20*(8), 1025–1041. https://doi.org/10.1016/j.acn.2005.05.003

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.

Iverson, G. L., Le Page, J., Koehler, B. E., Shojania, K., & Badii, M. (2007). Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. *The Clinical Neuropsychologist*, *21*(3), 532–546. https://doi.org/10.1080/13854040600611392

Jelicic, M., Ceunen, E., Peters, M. J., & Merckelbach, H. (2011). Detecting coached feigning using the Test of Memory Malingering (TOMM) and the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology*, *67*(9), 850–855. https://doi.org/10.1002/jclp.20805

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*(3), 410–425. https://doi.org/10.1076/clin.17.3.410.18089

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/10.1080/13854040701494987

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

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist*, *23*(5), 841–849. https://doi.org/10.1080/13854040902796735

Maiman, M., Del Bene, V. A., MacAllister, W. S., Sheldon, S., Farrell, E., Rentería, M. A., Slugh, M., Nadkarni, S. S., & Barr, W. B. (2018). Reliable Digit Span: Does it adequately measure suboptimal effort in an adult epilepsy population? *Archives of Clinical Neuropsychology*, *34*(2), 259–267. https://doi.org/10.1093/arclin/acy027

Menon, D. K., Schwab, K., Wright, D. W., & Maas, A. I. (2010). Position statement: Definition of traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *91*(11), 1637–1640. https://doi.org/10.1016/j.apmr.2010.05.0172010.05.017

Meulemans, T., Adam, S., & Palma-Duran, C. (2003). Cross-validation of the process-dissociation procedure and the Amsterdam Short-Term Memory Test for the detection of feigned memory deficits. Unpublished manuscript.

Meyers, J. E., & Meyers, K. R. (1995). Rey complex figure test under four different administration procedures. *The Clinical Neuropsychologist*, *9*(1), 63–67. https://doi.org/10.1080/13854049508402059

Meyers, J. E., & Volbrecht, M. (1998). Validation of reliable digits for detection of malingering. *Assessment*, *5*(3), 303–307. https://doi.org/10.1177/107319119800500309

Meyers, J. E., Morrison, A. L., & Miller, J. C. (2001). How low is too low, revisited: Sentence repetition and AVLT-recognition in the detection of malingering. *Applied Neuropsychology*, *8*(4), 234–241. https://doi.org/10.1207/S15324826AN0804_6

Millis, S. R. (2009). Methodological challenges in assessment of cognition following mild head injury: Response to Malojcic et al. 2008. *Journal of Neurotrauma*, *26*(12), 2409–2410. https://doi.org/10.1089/neu.2008.0530

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1094–1102. https://doi.org/10.1076/jcen.24.8.1094.8379

Moore, B. A., & Donders, J. (2004). Predictors of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury*, *18*(10), 975–984. https://doi.org/10.1080/02699050410001672350

Reedy, S. D., Boone, K. B., Cottingham, M. E., Glaser, D. F., Lu, P. H., Victor, T. L., Ziegler, E. A., Zeller, M. A., & Wright, M. J. (2013). Cross validation of the Lu and colleagues (2003) Rey-Osterrieth Complex Figure Test effort equation in a large known-group sample. *Archives of Clinical Neuropsychology*, *28*(1), 30–37. https://doi.org/10.1093/arclin/acs106

Rey, A. (1958). *Clinical assessment in psychology*. Presses Universitaires de Paris.

Reynolds, C. R. (1998). Common sense, clinicians, and actuarialism in the detection of malingering during head injury litigation. In C. R. Reynolds (Ed.), *Detection of malingering during head injury litigation* (pp. 261–286). Springer US.

Rogers, R., & Bender, S. D. (2020). *Clinical assessment of malingering and deception* (4th ed.). Guilford Publications.

Roor, J. J., Dandachi-FitzGerald, B., Peters, M., Knoop, H., & Ponds, R. (2022). Performance validity and outcome of cognitive behavior therapy in patients with chronic fatigue syndrome. *Journal of the International Neuropsychological Society*, *28*(5), 473–482. https://doi.org/10.1017/S1355617721000643

Schagen, S., Schmand, B., de Sterke, S., & Lindeboom, J. (1997). Amsterdam Short-Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology*, *19*(1), 43–51. https://doi.org/10.1080/01688639708403835

Schmand, B., Lindeboom, J., Merten, T., & Millis, S. R. (2005). *Amsterdam Short-Term Memory Test manual*. PITS. https://doi.org/10.1037/t12622-000

Schroeder, R. W., Martin, P. K., & Odland, A. P. (2016). Expert beliefs and practices regarding neuropsychological validity testing. *The Clinical Neuropsychologist*, *30*(4), 515–535. https://doi.org/10.1080/13854046.2016.1177118

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

Serafim, A., Saffi, F., Da Silva, T. G. B., De Almeida, C. V., Hokama, E., De Barros, D. M., & Dias, Á. M. (2015). Forensic neuropsychological assessment: A review of its scope. *Archives of Clinical Psychiatry*, *42*(2), 63–67. https://doi.org/10.1590/0101-60830000000049

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545–561. https://doi.org/10.1076/1385-4046(199911)13:04;1-y;ft545

Stulemeijer, M., Andriessen, T. M., Brauer, J. M., Vos, P. E., & Van der Werf, S. (2007). Cognitive performance after mild traumatic brain injury: The impact of poor effort on test results and its relation to distress, personality and litigation. *Brain Injury*, *21*(3), 309–318. https://doi.org/10.1080/02699050701209980

Suhr, J. A., & Barrash, J. (2007). Performance on standard attention, memory, and psychomotor speed tasks as indicators of malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 131–170). Oxford University Press.

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

Wechsler, D. (2011). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV)*. Pearson.

Wortzel, H. S., & Arciniegas, D. B. (2012). Treatment of post-traumatic cognitive impairments. *Current Treatment Options in Neurology*, *14*(5), 493–508. https://doi.org/10.1007/s11940-012-0193-6

Zasler, N. D., & Bigler, E. (2017). Medicolegal issues in traumatic brain injury. *Physical Medicine and Rehabilitation Clinics of North America*, *28*(2), 379–391. https://doi.org/10.1016/j.pmr.2016.12.012

Zenisek, R., Millis, S. R., Banks, S. J., & Miller, J. B. (2016). Prevalence of below-criterion Reliable Digit Span scores in a clinical sample of older adults. *Archives of Clinical Neuropsychology*, *31*(5), 426–433. https://doi.org/10.1093/arclin/acw025