



# The Myth of High False-Positive Rates on the Word Memory Test in Mild TBI

Laszlo A. Erdodi<sup>1</sup> · Paul Green<sup>2</sup> · Christina D. Sirianni<sup>1</sup> · Christopher A. Abeare<sup>1</sup>

Received: 5 May 2019 / Accepted: 17 May 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

This study was designed to replicate previous reports of elevated false-positive rates (FPR) on the Word Memory Test (WMT) in patients with mild traumatic brain injury (TBI) and to evaluate previous claims that genuine memory deficits and non-credible responding are conflated on the WMT. Data from a consecutive case sequence of 170 patients with mild TBI referred for neuropsychological assessment were collected. Failure rate on the WMT was compared to that on other performance validity tests (PVTs). The clinical characteristics and neuropsychological profiles of patients who passed and those who failed the WMT and other PVTs were compared. Base rate of failure was the highest on the WMT (44.7%), but comparable to that on other established PVTs (39.4–41.8%). The vast majority of patients (94.7%) who failed the WMT had independent evidence of invalid performance, refuting previous estimates of 20–30% FPR. Failing the WMT was associated with globally lower scores on tests measuring various cognitive domains. The neurocognitive profile of individuals with invalid performance was remarkably consistent across various PVTs. Previously reported FPR of the WMT were not replicated. Failing the WMT typically occurred in the context of failing other PVTs too. Results suggest a common factor behind non-credible responding that is invariant of the psychometric definition of invalid performance. Failure on the WMT should not be discounted based on rational arguments unsubstantiated by objective data. Inferring elevated FPR from high failure rate alone is a fundamental epistemological error.

**Keywords** Word Memory Test · False-positive rate · Performance validity · Mild TBI

## Introduction

### Performance Validity Assessment and the Word Memory Test (WMT)

An independent survey by Rabin et al. (2014) showed that the three most widely used free-standing performance validity tests (PVTs) were the TOMM (Test of Memory Malingering, Tombaugh, 1996), the Word Memory Test (WMT, Green & Astner, 1995; Green, 2003, Green, Lees-Haley, & Allen, 2003), and the MSVT (Medical Symptom Validity Test; Green, 2004). These tests were designed to be very easy and to provide an indication of whether the examinee produced a credible response set on neuropsychological testing. Failing

PVTs suggests that performance on ability tests cannot be treated as valid (Green, Rohling, Lees-Haley, & Allen, 2001; Boone, 2013; Larrabee, 2012).

The basic validation of the WMT began by applying the test to healthy volunteers, all of whom scored above the standard cutoffs, revealing perfect specificity (1.00; Green, Lees-Haley, & Allen, 2003). Out of 25 psychologists and physicians, who were asked to take the WMT and to simulate early dementia, 24 fell below the cutoffs. Hence, the sensitivity of the test to experimental malingering was .96. Very similar results were obtained by independent researchers comparing simulators and good effort volunteers in Turkish (Brockhaus & Peker, 2003), Russian (Tydecks, Merten, & Gubbay, 2006) and German (Brockhaus, & Merten, 2004).

In the German study, the WMT had perfect sensitivity (1.00) to simulated impairment, whereas only one of 32 institutionalized adults with intellectual disability failed the WMT (.97 specificity). In fact, the WMT had perfect specificity because it was later discovered that the woman who failed it had a borderline personality disorder and was malingering intellectual disability to obtain accommodation (R. Brockhaus, personal communication, 2010). Tydecks, Merten, and

✉ Laszlo A. Erdodi  
lerdodi@gmail.com

<sup>1</sup> Department of Psychology, University of Windsor, 168 Chrysler Hall South, 401 Sunset Ave, Windsor, ON N9B 3P4, Canada

<sup>2</sup> Green's Publishing, Kelowna, BC, Canada

Gubbay (2006) reported that the Russian WMT yielded perfect sensitivity and specificity in healthy Russian speaking volunteers assigned to a control or experimental malingering condition.

Native Turkish speakers living in Turkey or Germany were administered the Turkish WMT and asked to feign impairment (Brockhaus & Peker, 2003; Fritze, 2003). All 50 participants in the simulator group were detected (1.00 sensitivity) and 47 out of 48 volunteers in the control group passed the WMT (.98 specificity). The sole false positive was likely due to dyslexia or limited literacy, because the person was 67 years old, and reported zero years of formal education.

### The WMT in Children and Adults with Severe Dyslexia

Only 10% of a sample of children with severe dyslexia (Larochette & Harrison, 2012) failed the WMT, a population that is at high risk for failing a PVT that requires word reading. This is consistent with subsequent reports that severe dyslexia does not prevent credible examinees from passing PVTs based on the forced choice recognition paradigm that appear to require intact single word reading ability during the encoding trial (Hurtubise, Scavone, Sagar, & Erdodi, 2017). Following the recommendation to use the oral rather than the computerized WMT if the person's reading level is less than grade 3 may further decrease failure rate.

### False-Positive Rates (FPR): Facts and Controversy

The false-positive rate (FPR) and false-negative rate on PVTs are important issues. A false positive on a PVT may lead an assessor to misclassify a person with genuine impairment as non-credible. Conversely, a false negative on a PVT could result in concluding severe and debilitating cognitive impairment where none is actually present.

### A Myth Is Born: Failing the WMT = False-Positive Error

In head injury litigation, it is not unusual for the plaintiff expert to argue that failure on PVTs was caused by severe genuine impairment. The notion that the WMT recognition subtests actually measure ability and not effort, thereby substantiating the myth of high FPR, was proposed by Bowden, Shores, and Mathias (2006). They questioned the finding of a greater base rate of failure on the WMT in mild TBI than in severe TBI, as reported by Green, Rohling, Lees-Haley, and Allen (2001). Instead, they hypothesized that the WMT recognition subtests were sensitive to genuine deficits and, therefore, behaved like ability tests. They reported similar failure rates on the WMT across injury severity ranges.

This conclusion is problematic for a number of reasons. First, Green, Rohling, Lees-Haley, and Allen (2001) reported that WMT failure rates in *adults* with mild TBI are greater

than in *adults* with severe TBI. In contrast, Bowden, Shores, and Mathias (2006) included both children and adults with TBI in their sample. Notably, the percentage of children versus adults was not reported in the Bowden, Shores, and Mathias (2006) study—a methodologically relevant oversight that complicates the interpretation of their findings.

Second, the authors did not administer the WMT in the standard way, which involves multiple subtests, including Immediate Recognition (IR), Delayed Recognition (DR), Multiple Choice (MC), Paired Associate Recall (PA), Free Recall (FR), and Delayed Free Recall (LDFR). Instead, Bowden, Shores, and Mathias (2006) administered only the first subtest, IR. It is assumed that they used the oral version, which increases the risk of examiners inadvertently influencing the outcome of testing, as well as recording and scoring errors.

When Rohling and Demakis (2010) reanalyzed the raw data, they arrived at a different conclusion: the WMT did not behave like an ability test. The WMT-IR scores were uncorrelated with age and FSIQ, which supports the original designation of the WMT as a measure of performance validity, not cognitive ability. This finding was also consistent with other reports that WMT scores were independent of age and overall intellectual functioning (Lichtenstein, Erdodi, Rai, Mazur-Mosiewicz, & Flaro, 2018; Lichtenstein, Flaro, Baldwin, Rai, & Erdodi, 2019; Rohling & Demakis, 2010). In addition, the failure rate on the WMT-IR was higher (38%) in the mild TBI than in the moderate to severe TBI subsample (28.5%). If the WMT-IR subtests measured ability, the opposite would have been expected: higher  $BR_{Fail}$  in more severe TBI. Failure on the WMT is usually defined as a score of below the standard cutoff on IR, DR or the consistency score (CNS), not just the IR.

Far more people fail the WMT using the standard criteria than the IR subtest alone. In a sample of 2162 compensation seeking adults 643 failed the WMT (29.7%), but only 454 failed the IR (21.0%; R. Gervais, personal communication, 2019). By deviating from the standard administration protocol, Bowden, Shores, and Mathias (2006) undermined the psychometric integrity of the test, probably resulting in an artificially low failure rate, which in turn may have skewed the findings. If the WMT had been administered in the standard way (i.e., IR, DR, and CNS) to adults only, the findings may have been different.

Other authors who have argued that failures on the WMT in patients with mild TBI were false positives include Greve, Ord, Curtis, Bianchini, and Brennan (2008). They reported a 30% FPR on the WMT in a mixed sample including some mild TBI cases. In that study, non-credible responding was psychometrically defined based on passing or failing a certain set of PVTs, mainly the criteria used by Meyers, Volbrecht, Axelrod, and Reinsch-Boothby (2011). Thus, if patients passed these criteria but failed the WMT, they were classified

as false positives for the WMT. Notably, several of the validity cutoffs used to define invalid performance in this study were shown to be overly conservative (i.e., disproportionately sacrificing sensitivity for specificity) by subsequent research (Rai, 2019; Blaskewitz, Merten, & Brockhaus, 2009; Whiteside, Wald, & Busse, 2011).

More recently, Hall, Worthington, and Venables (2014) reported a failure rate of 18% in their acute mild TBI sample, and they equated this to the FPR on the WMT. Their argument was that these cases had passed a combination of other PVTs, including the TOMM (*Fail* defined as  $Trial\ 2 \leq 44$ ), Reliable Digit Span (*Fail* defined as  $\leq 6$ ), the Processing Speed Index of the WAIS-III (*Fail* defined as  $\leq 75$ ) and, therefore, demonstrated valid performance. They concluded that the mild TBI cases who failed the WMT were unable to pass because they had “verbal processing deficits.” Once again, these cutoffs have been shown to prioritize specificity at the expense of sensitivity (Erdodi et al., 2017a; Etherton, Bianchini, Heinly, & Greve, 2006; Jones, 2013; Kulas, Axelrod, & Rinaldi, 2014; Mathias et al., 2002; Reese, Suhr, & Riddle, 2012). As such, by virtue of selecting highly conservative cutoffs, the authors created a signal detection environment in which the WMT appears prone to high FPR by design.

### Rationally and Empirically Based Criticisms of the Myth of High FPR on the WMT

These conclusions are problematic on both rational and methodological grounds. Greve, Ord, Curtis, Bianchini, and Brennan (2008) provided no argument to support the assumption that mild TBI could result in cognitive impairment sufficient to cause failure on the WMT. An equally plausible alternative explanation is that those who passed Meyers’ criteria but failed the WMT were false negatives for the Meyers’ criteria. In fact, this is quite likely, because Meyers’ criteria are based on ability tests like Digit Span, which are correlated with age and intelligence, unlike the WMT recognition trials (Lichtenstein, Erdodi, Rai, Mazur-Mosiewicz, & Flaro, 2018; Lichtenstein, Flaro, Baldwin, Rai, & Erdodi, 2019; Rohling & Demakis, 2009). To reduce false positives on Meyers’ criteria, the cutoffs were set low, resulting in high specificity but low sensitivity. Optimizing the signal detection model for specificity is certainly a legitimate strategy, but its methodological implications should be made explicit (i.e., that creating a highly conservative criterion measure artificially inflates FPR on the WMT.)

Likewise, Hall, Worthington, and Venables (2014) failed to acknowledge that the WMT is more sensitive than the TOMM at standard cutoffs (Erdodi & Rai, 2017; Green, 2007; Green et al., 2004). Therefore, the discrepancies between PVTs may be better explained by the superior sensitivity of the WMT. More importantly, when failure on a PVT is observed, Slick et al. (1999) state that it is necessary to rule out the possibility

that failure on the PVT was caused by a genuine neurological condition. To do so, we need evidence that the condition in question (in this case mild TBI) involves sufficient impairment to explain the PVT failure.

No study has ever shown that mild TBI can produce impairment which is severe enough to cause failure on easy tests like the WMT, the MSVT, or the TOMM. In fact, a meta-analysis by Rohling et al. (2011) found no measurable impairment on standard measures of cognitive ability 3 months after mild TBI. This is consistent with research by McCrea (2008), which showed no detectable cognitive impairment 12 h after mild TBI in athletes who were tested both pre and post mild TBI on various neuropsychological tests. Therefore, persistent cognitive impairment is not expected after an uncomplicated mild TBI.

Similarly, there is no empirical support in the literature that people with mild TBI have acquired “verbal processing deficits” of any type, and certainly none sufficiently severe to cause WMT failure. In credible examinees, WMT failure is only seen in patients with dementia (Green, Montijo, & Brockhaus, 2011) or severe dyslexia (Larochette & Harrison, 2012). Most relevant to the present argument, WMT failure is rarely seen in people with moderate to severe TBI (Carone, 2008; Erdodi & Rai, 2017; Green, Lees-Haley, & Allen, 2003). In fact, the cutoffs for the WMT recognition measures were chosen specifically to be three standard deviations below the mean from patients with moderate to severe brain injury, who were assumed to demonstrate valid performance (Allen & Green, 1999).

Not surprisingly, the standard cutoffs had .98 specificity in the moderate to severe TBI group. In addition, the cutoffs were almost three standard deviations below the mean from mixed neurological outpatients who were disabled from work by medically verified neurological disorders (brain tumor, stroke, ruptured aneurysm, or other brain diseases; Green & Allen, 1999). As a result, the WMT cutoffs had a specificity of .93 in this population. It defies logic to argue that a mild TBI can cause WMT failure, when such failure implies scoring far worse than average for people with genuine cognitive deficits acquired subsequent to a severe TBI, and far worse than most people with disabling neurological diseases.

If a substance causes cancer, more of that substance should create more cancer, according to the Bradford Hill criteria for medical causation (Hill, 1965). In other words, there should be a “dose-response relationship.” Similarly, if we wished to argue that TBI causes failure on the WMT, we would have to show a greater failure rate in severe TBI than in mild TBI. However, the empirical data show just the opposite. Green, Rohling, Lees-Haley, and Allen (2001) reported that there was a 34% WMT failure rate in cases of mild TBI but only an 18% failure rate in those with moderate to severe TBI. An even greater discrepancy in WMT failure rates was reported by

Erdodi, Kirsch, Sabelli, and Abeare (2018) between patients with mild (46.1%) and moderate to severe (16.7%) TBI.

These findings extended to the MSVT and NV-MSVT as well as several other PVTs. In all cases but one (Warrington's Recognition Memory Test – Words; Warrington, 1984), the failure rate was significantly higher in mild TBI than in severe TBI (Green & Merten, 2013). This reverse dose-response relationship argues beyond reasonable doubt that brain injury cannot explain WMT, MSVT, or NV-MSVT failures.

Another source of convincing evidence comes from adults and children diagnosed with intellectual disability, who were nevertheless able to pass the WMT (Brockhaus & Merten, 2004; Green, Flaro, Brockhaus, & Montijo, 2012; Green & Flaro, 2016). Even patients with left hippocampotomy (Carone, Green, & Drane, 2013) or bilateral hippocampal damage have been reported to pass the WMT (Goodrich-Hunsaker & Hopkins, 2009). A person who fails the WMT is performing in the same range as patients with dementia (Green, Montijo, & Brockhaus, 2011). The literature on mild TBI is extensive and it does not support dementia level impairment in people with mild TBI.

Carone (2014) showed that a 9-year-old girl with intractable seizures, massive developmental absence of brain tissue from hydrocephalus, and FSIQ < 60 produced near-perfect scores on the WMT and MSVT recognition trials and thus, could easily pass these PVTs. Therefore, it would be highly unusual for an adult with only a mild TBI and credible responding to fail these PVTs. Previous research suggests that failure on the WMT cannot be attributed to depression (Rohling, Green, Allen, & Iverson, 2002; Williamson, Holsman, Chaytor, Miller, & Drane, 2012), PTSD (Demakis, Gervais, & Rohling, 2008), or fibromyalgia (Gervais et al., 2001; Suhr, 2003). Although some investigators found that patients with comorbid TBI and PTSD had an increased cumulative failure rate using multiple PVTs (Clark, Amick, Fortier, Milberg, & McGlinchey, 2014; Greiffenstein & Baker, 2008), the clinical interpretation of the role psychogenic factors play in PVT failure is an evolving debate (Henry et al., 2018). Regardless of underlying etiology, non-credible responding is the most parsimonious explanation of WMT failure in mild TBI (Delis & Wetter, 2007; Erdodi, Abeare et al., 2018).

PVT failure rate is especially high among individuals with mild TBI who are involved in compensation seeking. Mittenberg, Patton, Canyock, and Condit (2002) estimated from multiple sources that 39% of people with mild TBI produce invalid data in a compensation context. Larrabee's (2009) survey concluded that PVT failures occurred in approximately 40% of cases involved with litigation or compensation, including people with mild TBI. The single largest group surveyed by Larrabee (2009) contained 719 cases taken from the practice of the second author of the present paper. Of this sample, 40% failed the WMT or the Computerized

Assessment of Response Bias (CARB; Conder, Allen, & Cox, 1992). Larrabee, Millis, & Meyers, (2009) used the term "the magic number of 40% plus or minus 10" to describe the incidence of PVT failure among individuals evaluated in the context of external incentives to appear impaired.

### Comparison to Other Free-Standing PVTs

The MSVT was deliberately designed to be even easier than the WMT, and has been shown to be easily passed by children with severe TBI (Carone, Green, & Drane, 2013) as well as by adults with acute severe TBI, once they emerge from post-traumatic amnesia (Macciocchi, Seel, Yi, & Small, 2017). As long as they had grade 3 reading levels, all children with intellectual disability passed the MSVT (Green & Flaro, 2015). In fact, the MSVT is so easy that when children who spoke no French were tested in French, they scored a mean of 98% correct on the recognition subtests, which was the same as fluent French speaking adults (Richman et al., 2006). This finding demonstrates that, unlike other free-standing PVTs based on forced choice word recognition like the Word Choice Test (Pearson, 2009), the MSVT is unaffected by limited English proficiency (Erdodi, Nussbaum, Sagar, Abeare, & Schwartz, 2017).

In a large sample of adults with mild TBI, Green, Flaro, and Courtney (2009) used performance on the MSVT to evaluate the possibility that WMT failures were false positives. Those who failed the WMT were, in most cases, also administered the even easier MSVT. If true deficits accounted for the WMT failure, one would expect a lower failure rate on the easier test, the MSVT. However, this was not the case. The majority of patients who failed the WMT also failed the MSVT. In fact, the group average was significantly lower on the MSVT than a group of children with intellectual disabilities. There is nothing in the mild TBI literature to suggest that such severe intellectual impairment can be acquired as a result of mild TBI. In addition, the mild TBI cases who failed the WMT also scored significantly lower on the MSVT than a group of children with developmental disability, who had been specifically selected for having impaired verbal memory. The literature on mild TBI in adults does not provide any support for the link between mild TBI and impaired verbal memory. Thus, it was concluded that patients with mild TBI who failed the WMT were true positives.

### The Limits of Legitimate Exemptions from PVT Failures

Admittedly, some patients do suffer from cognitive impairment severe enough to cause failure on PVTs. For example, 27% of patients with dementia described on pages 43 to 45 of the TOMM test manual failed the TOMM Trial 2 (Tombaugh, 1996). Similarly, people with dementia, mainly of an

Alzheimer's type, failed the WMT or the MSVT (Green, Montijo, & Brockhaus, 2011). It appears that, whereas most forms of brain disease do not cause PVT failure, some credible examinees will fail PVTs if their impairment is severe enough, such as Alzheimer's disease.

The problem with using one PVT or a combination of PVTs to validate another PVT as in the studies of Greve, Ord, Curtis, Bianchini, and Brennan (2008) and Hall, Worthington, and Venables (2014) is that it introduces circular reasoning (Bigler, 2015). For example, in a series of 1500 patients tested by R. Gervais (personal communication, 2016), the TOMM and Reliable Digit Span (RDS) were used to define invalid test data. Only 6% of compensation seeking cases produced invalid results (i.e., failed both the TOMM and the RDS). Yet in the same group, 30% of cases failed the WMT. In addition, the WMT was not administered if examinees had dyslexia or less than a grade 3 reading level to further protect against an elevated FPR.

One possible conclusion is that the WMT produced a FPR of 24%, consisting mainly of patients with orthopedic injuries, depression, anxiety, PTSD, and dealing with situational stressors. However, it would be hard to argue that an adult with normal premorbid cognitive abilities was unable to pass the TOMM, the WMT or the MSVT due to a mild TBI. To maintain that the high failure rate in the WMT reflects elevated FPR, one would have to adopt the indefensible argument that these examinees assessed in outpatient setting failed the WMT because of genuine and severe cognitive impairment equivalent to that seen in some dementias. No other neurological condition has been shown to reliably elevate failure rates on the WMT.

### The Current Study: Rationale, Design, and Hypothesis

The above summary suggests that the WMT has high specificity in both children and adults with genuine cognitive impairment, with the notable exceptions of dementia, and severe dyslexia. In the current study, we explored further the issue of FPR on the WMT in adults with mild TBI. We selected a large case series of patients with mild TBI who were administered the WMT in conjunction with other PVTs—both stand-alone and embedded. We hypothesized that the actual FPR on the WMT will be low (i.e.,  $\leq 10\%$ ). In order to deem a WMT failure a *false positive*, it must occur in the context of psychometric evidence of credible overall responding. In other words, the examinee must fail no more than one other PVT. In contrast, if an examinee failed  $\geq 2$  independent PVTs in addition to the WMT, the failure on the WMT was considered a *true positive*, following established clinical and forensic guidelines (Boone, 2013; Larrabee, 2014). To circumvent some of the methodological flaws of previous studies, traditional free-standing PVTs were complemented by embedded validity indicators aggregated into a single validity composite, in order to harness the advantage of multivariate models of

performance validity assessment (An, Kaploun, Erdodi, & Abeare, 2017; Abeare et al., 2018; Cottingham, Victor, Boone, Ziegler, & Zeller, 2014; Erdodi et al., 2017a; Pearson, 2009; Tyson et al., 2018).

## Method

### Participants

Archival data were collected from a consecutive case sequence of 170 patients with mild TBI referred for neuropsychological assessment in a private practice setting in the setting of a compensation or disability claim. The majority were male (67.7%) and right-handed (87.4%). Mean age was 42.5 years ( $SD = 11.4$ ), while mean level of education was 12.2 years ( $SD = 2.4$ ). The most common reasons and sources for referral were the Workman's Compensation Board (42.5%), followed by personal injury lawyers (22.8%), independent medical examinations (15.6%), and clinical (10.8%). Inclusion criteria were claimed history of mild TBI [Glasgow Coma Scale (GCS)  $\geq 13$ , loss of consciousness  $< 30$  min, post-traumatic amnesia  $< 1$  h], being in the post-acute stage of recovery (i.e., assessed  $> 3$  months after the mild TBI), and a complete administration of the WMT and MSVT. Rai & Erdodi 2019.

### Materials

A core battery of neuropsychological tests assessing attention, memory, processing speed, and executive functioning was administered to all patients, in addition to the two free-standing PVTs. Given the similarity between the WMT and MSVT (both free-standing computerized tests with recognition subtests based on the forced choice recognition paradigm) and previous research suggesting that shared features between the predictor and criterion variable can influence the outcome of classification accuracy analyses (Erdodi, 2019), two additional validity composites were developed by combining data from multiple embedded PVTs. The purpose of creating these aggregate measures of performance validity was to avoid the perception of bias in the choice of criterion PVT (i.e., evaluating the WMT against another PVT that is very similar in order to artificially inflate the concordance rate) (Rai & Erdodi, 2019).

The first composite was labeled "Validity Index Seven" (VI-7). As the name suggests, the VI-7 was built from seven independent embedded PVTs, some of which contained multiple indicators (Table 1). The VI-7 represents the traditional approach to multivariate models of performance validity assessment. Its value reflects the number of failures on the component PVTs. As such, it can range from zero (no failure) to seven (failure on all constituent PVTs). Each additional unit

**Table 1** The components of the VI-7, cutoffs, and base rates of failure

Test	Indicator	Scale	Cutoff	% ADM	BR <sub>Fail</sub>	Reference
BCT	TE	Raw	≥ 100	90.6	11.2	DeFilippis & McCampbell, 1979, 1991, 1997; Tenhula & Sweet, 1996; DiCarlo, Gfeller, & Oliveri, 2000
CVLT	Trials 1–5	Raw	≤ 37	91.2	21.2	Coleman et al., 1998; Curtis, Greve, Bianchini, & Brennan, 2006; Millis et al., 1995; Slick, Iverson, & Green, 2000; Shura et al., 2016; Sweet et al., 2000; Trueblood, 1994
		RecHITs	Raw	≤ 11		Coleman et al., 1998; Curtis, et al., 2006; Millis et al., 1995; Persinger et al., 2018; Shura et al., 2016; Slick et al., 2000; Sweet et al., 2000; Trueblood, 1994; Wolfe et al., 2010
GPB	DH	T	≤ 29	90.0	14.3	Erdodi, Seke et al., 2017; Erdodi, Kirsch, Sabelli, & Abeare, 2018
	ND	T	≤ 29			
	COM	T	≤ 31			
Digit Span	RDS	Raw	≤ 7	80.0	21.2	Greiffenstein, Baker, & Gola, 1994; Mathias et al., 2002; Pearson, 2009; Reese, Suhr, & Riddle, 2012
TMT	A	T	≤ 37	97.1	31.2	Abeare et al., 2019; Ashendorf, Clark, & Sugarman, 2017; Erdodi & Lichtenstein, 2019
TMT	B	T	≤ 35	96.5	25.9	Ashendorf, Clark, & Sugarman, 2017; Erdodi & Lichtenstein, 2019
WCST	FMS	Raw	≥ 2	76.5	30.0	Greve, Heinly, Bianchini, & Love, 2009; Gligorovic & Buha, 2013; Larrabee, 2003; King et al., 2002; Suhr & Boyer, 1999

VI-7, Validity Index Seven; BCT, Booklet Category Test; CVLT, California Verbal Learning Test; GPB, Grooved Pegboard Test; TMT, Trail Making Test; WCST, Wisconsin Card Sorting Test; TE, total errors; RecHITs, Recognition hits (number of target words correctly recognized out of 16); DH, dominant hand; ND, non-dominant hand; COM, combined (Pass defined as one or both hands scoring > 31; Fail defined as both hands scoring ≤ 31); RDS, reliable digit span; FMS, failure to maintain set; T, Demographically adjusted score using the norms by Heaton, Miller, Taylor and Grant (2004); %<sup>ADM</sup>, percent of the sample to which the test was administered; BR<sub>Fail</sub>, base rate of failure

increase represents an increased likelihood that the overall response set is invalid (Table 2). However, for the purpose of classification accuracy analyses, an overall *Pass* on the VI-7 was defined as ≤ 1 (at most one PVT failure), and *Fail* was defined as ≥ 2 (at least two PVT failures), following commonly accepted forensic standards for embedded measures (Boone, 2013; Larrabee, 2014).

The second composite was labeled “Erdodi Index Five” (EI-5). The EI-5 represents a novel approach to aggregating PVTs designed to recapture the underlying continuity in performance validity (Erdodi et al., 2017a) by simultaneously measuring both the number and extent of PVT failures (Erdodi, Hurtubise et al., 2018). Instead of dichotomizing

(Pass/Fail) its components, the EI-5 model recodes each constituent PVT onto a four-point ordinal scale, where zero means an unequivocal Pass, and three means an unequivocal Fail, with two intermediate levels of failure (Table 3). An EI-5 value of one is defined by the most liberal cutoff available in the literature. As such, it is optimized for sensitivity. Conversely, an EI-5 value of three is defined by the most conservative cutoff available in the literature or in the absence of a natural threshold, the bottom 5% of the distribution on an embedded PVT nested in a measure of cognitive ability. An EI-5 value of two represents an intermediate level of failure, and is therefore associated with a cutoff more conservative than an EI-5 value of one, but more liberal than an EI-5 value

**Table 2** The distribution of scores for the VI-7 and clinical classification ranges

VI-7	<i>f</i>	%	Cumulative %	Classification
0	54	31.8	31.8	PASS
1	45	26.5	58.2	Pass
2	29	17.1	75.3	Fail
3	19	11.2	86.5	FAIL
4	15	8.8	95.3	<b>FAIL</b>
5	6	3.5	98.8	<b>FAIL</b>
6	2	1.2	100.0	<b>FAIL</b>

VI-7, Validity Index Seven; darker shading indicates stronger cumulative evidence of invalid performance

of three. In the absence of a natural demarcation line, it is defined as “the next 5% of worst performance” following the bottom 5% of the distribution. As such, its upper limit is roughly defined by the 10th percentile, while its lower limit is defined by the 5th percentile (Erdodi, 2019).

The value of the EI-5 is obtained by summing its recoded components. Therefore, it can range from zero (the examinee passed the most liberal cutoff on all five components) to 15 (the examinee failed the most conservative cutoff on all five components). An  $EI-5 \leq 1$  reflects at most one marginal failure, and is consequently labeled an overall Pass. The next two levels (2–3) are problematic, as they could represent either several independent failures at the most liberal cutoff or a single failure at a more conservative cutoff. Although in previous studies, this level of performance has been associated with significantly stronger independent evidence of invalid performance compared to the Pass range (Erdodi, 2019; Erdodi, Hurtubise et al., 2018; Erdodi, Seke et al., 2017), it does not meet the strict definition of global non-credible responding, and it has a relatively high multivariate base rate (Pearson, 2009). As such, this range was labeled Borderline, and excluded from analyses that require a dichotomous (Pass/Fail) criterion measure.

However, an  $EI-5 \geq 4$  crosses the “red line”: it represents either a minimum of four independent failures at the most liberal cutoff, two at the more conservative cutoff, or one at the most liberal and one at the most conservative. Regardless

of the specific configuration, this combination of PVT failures meets the commonly accepted definition of invalid performance and has a low multivariate base rate (Pearson, 2009). Therefore, this range of EI-5 was labeled a *Fail*, with increasing confidence in the classification with each unit increase in the EI-5 value (Erdodi, Dunn et al., 2018), as shown in Table 4.

To illustrate the shift in signal detection profile as a function of cutoffs, classification accuracy was computed against the MSVT across a range of cutoffs. When failure on the EI-5 was defined as  $\geq 1$ , specificity was unacceptably low (.50). Increasing the cutoff to  $\geq 2$  ( $Pass \leq 1$ ) only slightly improved specificity (.63). Further increasing the cutoff to  $\geq 3$  produced steady, but still insufficient improvement in specificity (.77). However, the standard  $\geq 4$  cutoff resulted in a good combination of sensitivity (.57) and specificity (.93). It also proved to be the point of diminishing returns: making the cutoff any more conservative produced marginal improvement in specificity while disproportionately sacrificing sensitivity.

Given that the EI-5 is a relatively new method of aggregating independent PVTs, its classification was first validated against established free-standing instruments. The EI-5 was a significant predictor of passing or failing the MSVT:  $AUC = .75$  (95% CI .66–.82), with .57 sensitivity and .93 specificity. The EI-5 produced a similar

**Table 3** Individual components of the EI-5 and base rates of failure at given cutoffs

EI-5 Components	EI-5 Values			
	0	1	2	3
	Pass	Fail	FAIL	FAIL
CVLT RecHITs	>11	11	10	≤9
<i>Base Rate</i>	89.0%	1.9%	3.2%	5.8%
FMS <sub>WCST</sub>	<2	2	3	≤4
<i>Base Rate</i>	60.8%	11.5%	16.9%	10.8%
RDS	>7	7	6	≤5
<i>Base Rate</i>	72.8%	16.9%	5.1%	5.1%
TMT-A	≥37	29-37	21-28	≤20
<i>Base Rate</i>	67.9%	21.2%	6.1%	4.8%
TMT-B	≥35	27-35	24-26	≤23
<i>Base Rate</i>	73.2%	16.4%	4.8%	5.5%

*EI-5*, Erdodi Index Five; *CVLT RecHITs*, recognition hits on the California Verbal Learning Test (raw score; Coleman et al., 1998; Curtis, Greve, Bianchini, & Brennan, 2006; Millis et al., 1995; Persinger et al., 2018; Shura et al., 2016; Slick et al., 2000; Sweet et al., 2000; Trueblood, 1994; Wolfe et al., 2010); *RDS*, reliable digit span (Greiffenstein, Baker, & Gola, 1994; Mathias et al., 2002; Pearson, 2009; FMS<sub>WCST</sub>, failure to maintain set on the Wisconsin Card Sorting Test Greve, Heinly, Bianchini, & Love, 2009; Gligorovic & Buha, 2013; Larrabee, 2003; King et al., 2002; Suhr & Boyer, 1999; *TMT-A*, Trail Making Test (*T* score; Ashendorf, Clark, & Sugarman, 2017); darker shading indicates stronger cumulative evidence of invalid performance

**Table 4** Frequency, cumulative frequency and classification range for the first ten levels of the EI-5

EI-5	<i>f</i>	%	<i>Cumulative%</i>	MSVT		Classification	
				SENS	SPEC	By Row	Overall
0	63	37.1	37.1	-	-	PASS	PASS
1	22	12.9	50.0	.82	.50	Pass	
2	31	18.2	68.2	.70	.63	Borderline	
3	23	13.5	81.8	.64	.77	Borderline	FAIL
4	13	7.6	89.4	.57	.93	Fail	
5	7	4.1	93.5	.43	.96	Fail	
6	3	1.8	95.3	.33	.98	FAIL	
7	3	1.8	97.1	.26	.98	FAIL	
8	1	0.6	97.6	.17	.98	<b>FAIL</b>	
9	1	0.6	98.2	.13	.98	<b>FAIL</b>	

EI-5, Erdodi Index Five; MSVT, Medical Symptom Validity Test (standard cutoffs); darker shading indicates stronger cumulative evidence of invalid performance

signal detection profile against the Non-Verbal MSVT: AUC = .71 (95% CI .59–.83), .41 sensitivity, and .87 specificity.

## Procedure

Psychometric testing was administered and scored by trained technicians in an outpatient setting. A licensed clinical neuropsychologist performed the clinical interview, reviewed the available records, prepared an integrative report, and rendered the final diagnosis. Patients gave written consent for anonymized retrospective analysis of their test results in group data. Only de-identified data were captured for research purposes to protect patient confidentiality. APA ethical guidelines regulating research involving human participants were followed throughout the process.

## Data Analysis

Basic descriptive statistics (M, SD, frequency distribution,  $BR_{Fail}$ ) were reported where relevant. The main inferential statistics were independent *t* tests for continuous variables and  $\chi^2$  test of independence or two proportions *z* tests for categorical variables. The statistical significance of the difference between SDs was evaluated using Levene's test of homogeneity of variances. Effect size estimates were given in Cohen's *d* and Cramer's  $\phi^2$ . Overall classification accuracy (AUC) and the associated 95% CI were calculated using SPSS 23.0. Sensitivity and specificity were computed using standard formulae (Grimes & Schultz, 2005).

## Results

### Base Rates of Failure Across PVTs

The highest failure rate was observed on the WMT (44.7%), followed by the VI-7 (41.8%), the MSVT (39.4%), and the EI-5 (18.2%). The difference in failure rate between the WMT and VI-7 or the MSVT was not statistically significant ( $p = .322$  and  $.584$ , respectively). However, patients were significantly ( $p < .001$ ) more likely to fail the WMT compared to the EI-5 (risk ratio 2.46).

### The Rate of Agreement Between WMT and Criterion PVTs

The dichotomous outcome (Pass/Fail) of the WMT and that of the MSVT were strongly related:  $\chi^2(1) = 78.4$ ,  $p < .001$ ,  $\phi^2 = .46$  (very large effect). The two tests agreed on classifying 84.1% of the sample. The relationship with the VI-7 was weaker, but still highly significant:  $\chi^2(1) = 40.2$ ,  $p < .001$ ,  $\phi^2 = .24$  (very large effect). The two tests agreed on classifying 74.7% of the sample. Finally, there was a large proportion of shared variance with the EI-5:  $\chi^2(1) = 30.1$ ,  $p < .001$ ,  $\phi^2 = .26$  (very large effect). The two tests agreed on classifying 75.9% of the sample.

### Clinical Characteristics and Neurocognitive Profiles Associated with Failing the WMT

Patients who failed the WMT were significantly older and reported higher level of depression (medium effect) than those who passed the WMT. No difference was found on gender, education, GCS, and lateral dominance (Table 5). Failing the



WMT was associated with significantly lower performance on measures of auditory attention, learning and memory (large effects), concept formation (large effects), visuomotor processing speed (large effects), and manual dexterity (large effects). In addition, the patients who failed the WMT produced significantly more variable scores on a test of concept formation and a derivative index of executive deficits.

### Clinical Characteristics and Neurocognitive Profiles Associated with Failing the VI-7

Patients who failed the VI-7 were significantly older and reported higher levels of depression (medium effect). No difference was found on gender, education, GCS, lateral dominance, and manual dexterity (Table 6). Failing the VI-7 was associated with significantly lower performance on measures of auditory attention, learning and memory (large effects), concept formation (large effects), and visuomotor processing speed (large effects). As observed with the WMT, patients

who failed the VI-7 produced significantly more variable scores on a test of concept formation and a derivative index of executive deficits.

### Clinical Characteristics and Neurocognitive Profiles Associated with Failing the EI-5

Patients who failed the EI-5 reported significantly higher levels of depression (medium effect). No difference was found on age, gender, education, GCS, and lateral dominance (Table 7). Failing the EI-5 was associated with significantly lower performance on measures of auditory attention (medium effect), auditory learning and memory (large effects), concept formation (large effects), visuomotor processing speed (large effects), and manual dexterity (large effect). Patients who failed the EI-5 produced significantly more variable scores during the delayed free recall trial of an auditory verbal learning test and a measure of visual attention, sequencing, and motor speed.

**Table 5** Comparing patients who passed and those who failed the Word Memory Test on demographic variables and performance on neuropsychological tests

		Word Memory Test						<i>t</i>	<i>p</i>	<i>d</i>	$\sigma_1$ vs. $\sigma_2$
		Pass			Fail						
		<i>n</i>	M	SD	<i>n</i>	M	SD				
CVLT	Age	94	40.0	11.2	76	45.6	10.8	3.29	.001	.51	.856
	Education	94	12.6	2.1	76	11.9	2.8	1.86	.064	–	.004
	GCS	63	14.7	0.6	50	14.7	0.6	0.09	.929	–	.980
	BDI-II	92	15.8	9.8	71	22.0	11.6	3.74	< .001	.58	.224
	% male*	94	34.0%		76	30.3%		0.27	.600	–	–
	% RH*	92	89.1%		74	85.1%		0.59	.441	–	–
	Trials 1–5	86	56.3	10.9	69	42.0	11.5	7.93	< .001	1.28	.943
	Trial 1	86	7.7	2.5	69	5.9	2.3	4.62	< .001	.75	.274
Trial 5	86	13.2	2.3	69	9.8	2.8	8.62	< .001	1.33	.114	
SD-FR	86	11.9	2.8	69	7.3	3.1	9.69	< .001	1.56	.510	
LDFR	86	12.2	2.8	69	7.8	3.2	9.07	< .001	1.46	.479	
WCST	CATcomp	87	5.5	1.3	65	3.6	2.1	6.80	< .001	1.43	< .001
BCT	TE	90	45.1	10.5	64	37.3	10.3	4.55	< .001	.75	.830
WAIS-R	CD	54	9.7	2.6	38	8.0	2.5	3.08	.003	.67	.716
TMT	A	91	48.0	10.7	74	39.1	13.6	4.67	< .001	.73	.062
	B	91	47.5	11.2	73	39.5	13.5	4.17	< .001	.64	.149
FTT	DH	19	50.6	10.3	10	36.1	14.5	3.13	.004	1.15	.478

GCS, Glasgow Coma Scale; BDI-II, Beck Depression Inventory – Second Edition; RH, right-handed; CVLT, California Verbal Learning Test (raw scores); WCST, Wisconsin Card Sorting Test; BCT, Booklet Category Test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); WAIS-R, Wechsler Adult Intelligence Scale – Revised; TMT, Trail Making Test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); FTT, Finger tapping test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); SD-FR, short delay free recall; LDFR, long delay free recall; CATcomp, categories completed; TE, total errors; CD, Digit-Symbol Coding (age-corrected scaled score); DH, dominant hand;  $\sigma_1$  vs.  $\sigma_2$ , the significance level (*p* value) associated with Levene's test of equality of variances

\*Chi-square test of independence

**Table 6** Comparing patients who passed and those who failed the VI-7 on demographic variables and performance on neuropsychological tests

		VI-7						<i>t</i>	<i>p</i>	<i>d</i>	$\sigma_1$ vs. $\sigma_2$
		Pass			Fail						
		<i>n</i>	M	SD	<i>n</i>	M	SD				
CVLT	Age	99	40.3	11.3	71	45.6	10.9	3.06	.003	.48	.892
	Education	99	12.4	2.3	71	12.1	2.7	0.71	.481	–	.060
	GCS	70	14.8	0.5	43	14.7	0.7	0.86	.393	–	.060
	BDI-II	95	16.3	10.4	68	21.6	11.2	3.15	.002	.49	.816
	% male*	99	35.4%		71	28.2%		0.98	.323	–	–
	% RH*	95	87.4%		71	87.3%		0.00	.993	–	–
	Trials 1–5	93	55.2	11.2	62	41.5	11.4	7.59	< .001	1.24	.962
	Trial 1	93	7.6	2.6	62	5.8	2.1	4.54	< .001	.76	.128
	Trial 5	93	13.0	2.5	62	9.8	2.9	7.11	< .001	1.18	.188
	SD-FR	93	11.4	3.3	62	7.5	3.2	7.27	< .001	1.20	.972
	LDFR	93	11.7	3.1	62	8.0	3.4	7.09	< .001	1.14	.677
WCST	CAT <sub>comp</sub>	87	5.5	1.2	65	3.5	2.1	7.18	< .001	1.17	< .001
BCT	TE	94	44.7	10.6	60	37.3	10.3	4.31	< .001	.71	.997
WAIS-R	CD	58	9.7	2.7	34	7.8	2.2	3.52	.001	.77	.305
FTT	DH	20	47.7	12.9	9	40.9	14.7	1.26	.217	–	.811

VI-7, Validity Index Seven (*Pass* defined as  $\leq 1$ ; *Fail* defined as  $\geq 2$ ); GCS, Glasgow Coma Scale; BDI-II, Beck Depression Inventory – Second Edition; RH, right-handed; CVLT, California Verbal Learning Test (raw scores); WCST, Wisconsin Card Sorting Test; BCT, Booklet Category Test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); WAIS-R, Wechsler Adult Intelligence Scale – Revised; TMT, Trail Making Test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); FTT, finger tapping test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); SD-FR, short delay free recall; LDFR, long delay free recall; CAT<sub>comp</sub>, categories completed; TE, total errors; CD, Digit-Symbol Coding (age-corrected scaled score); DH, dominant hand;  $\sigma_1$  vs.  $\sigma_2$ , the significance level (*p* value) associated with Levene's test of equality of variances

\*Chi-square test of independence

### Examining the False-Positive Rate on the WMT

Of the 76 patients who failed the WMT (44.7% of the sample), 72 also failed either another free-standing PVT or at least two embedded PVTs. In other words, 94.7% of patients identified as non-credible by the WMT also had independent evidence of invalid performance. This is equivalent to the overall specificity of the WMT relative to the other PVTs employed. Conversely, the false-positive rate on the WMT (i.e., proportion of the sample that failed the WMT, but had no other psychometric evidence of non-credible responding) was 5.3%. The vast majority of patients (92.1%) who failed the WMT also failed  $\geq 2$  other PVTs, 71.1% failed  $\geq 3$ , and 51.3% failed  $\geq 4$ .

### Discussion

This archival study empirically evaluated earlier claims that the WMT produces an unacceptably high FPR in patients with mTBI (Bowden, Shores, & Mathias, 2006; Hall, Worthington, & Venables, 2014; Greve, Ord, Curtis, Bianchini, & Brennan, 2008). The investigation was extended to four levels of analysis. First, the failure rate on the WMT was compared to that observed

on other PVTs. Second, the performance of patients who passed versus those who failed the WMT was compared across commonly used neuropsychological tests measuring core cognitive domains (attention and processing speed, memory, concept formation, mental flexibility, visuomotor scanning, and manual dexterity). Neurocognitive profiles, demographic variables, and injury parameters associated with valid vs. invalid performance on the WMT were then compared to those obtained using different criterion PVTs to examine the effect of the changing psychometric definitions of non-credible responding. Third, the concordance rate between the Pass/Fail outcome on the WMT and the other criterion PVTs was computed. Finally, an ipsative analysis was performed on the subset of patients who failed the WMT to determine the proportion of these cases with independent evidence of invalid performance.

Indeed, the highest failure rate (44.7%) was observed on the WMT, followed by the VI-7, a multivariate composite of embedded validity indicators (41.8%), the MSVT (39.4%) and the EI-5, a novel approach to aggregating multiple PVTs (26.7%). The lower failure rate on the EI-5 can be attributed to the conservative approach guiding its design: borderline range scores (2 and 3) are excluded from analyses that require a dichotomous outcome (Erdodi, 2019), even though a

**Table 7** Comparing patients who passed and those who failed the EI-5 on demographic variables and performance on neuropsychological tests

		EI-5						<i>t</i>	<i>p</i>	<i>d</i>	$\sigma_1$ vs. $\sigma_2$
		Pass			Fail						
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>				
CVLT	Age	85	41.6	12.1	31	45.0	10.9	1.38	.169	–	.760
	Education	85	12.3	2.4	31	12.9	2.9	1.11	.270	–	–
	GCS	63	14.8	0.5	19	14.7	0.7	0.46	.693	–	.359
	BDI-II	82	15.8	9.7	30	20.6	11.8	2.20	.030	.44	.230
	% male*	85	32.9%		31	38.7%		0.34	.563	–	–
	% RH*	82	87.8		31	83.8		0.30	.582	–	–
	Trials 1–5	79	53.9	11.6	26	39.8	12.8	5.26	< .001	1.15	.602
	Trial 1	79	7.4	2.6	26	5.9	2.3	2.62	.011	.61	.330
	Trial 5	79	12.6	2.8	26	9.2	3.5	5.01	< .001	1.08	.173
	SD-FR	79	10.9	3.4	26	7.4	3.9	4.51	< .001	.96	.373
	LDFR	79	11.5	3.1	26	8.0	4.3	4.54	< .001	.93	.028
WCST	CATcomp	75	5.4	1.5	25	3.2	1.9	6.09	< .001	1.29	.013
BCT	TE	81	43.6	10.3	23	33.9	11.4	3.90	< .001	.89	.769
WAIS-R	CD	49	10.0	2.7	12	6.8	1.8	3.95	< .001	1.39	.234
FTT	DH	20	48.5	12.4	5	35.6	14.7	2.01	.056	.95	.670

*EI-5*, Erdodi Index Five (*Pass* defined as  $\leq 1$ ; *Fail* defined as  $\geq 4$ ); *GCS*, Glasgow Coma Scale; *BDI-II*, Beck Depression Inventory – Second Edition; *RH*, right-handed; *CVLT*, California Verbal Learning Test (raw scores); *WCST*, Wisconsin Card Sorting Test; *BCT*, Booklet Category Test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); *WAIS-R*, Wechsler Adult Intelligence Scale – Revised; *TMT*, Trail Making Test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); *FTT*, finger tapping test (demographically adjusted *T* score based on norms by Heaton, Miller, Taylor and Grant, 2004); *SD-FR*, short delay free recall; *LDFR*, long delay free recall; *CATcomp*, categories completed; *TE*, total errors; *CD*, Digit-Symbol Coding (age-corrected scaled score); *DH*, dominant hand;  $\sigma_1$  vs.  $\sigma_2$ , the significance level (*p* value) associated with Levene's test of equality of variances

\*Chi-square test of independence

performance in this range was equally likely to indicate credible and non-credible responding based on the present findings (Erdodi, Hurtubise et al., 2018). Moreover, previous research suggests that EI-5 scores in the borderline range are *more likely to reflect invalid than valid performance* (Erdodi et al., 2018a; Erdodi et al., 2017b). As such, the EI-5 tends to underestimate the base rate of failure by design. Although the failure rate on the WMT within this sample was high, it is comparable to previous estimates of failure rates on PVTs in this population tested in both clinical (Abeare, Messa, Zuccato, Merker, & Erdodi, 2018; Erdodi, Kirsch, Sabelli, & Abeare, 2018; Erdodi & Roth, 2017) and forensic settings (Mittenberg, Patton, Canyock, & Condit, 2002; Larrabee, Millis, & Meyers, 2009).

Consistent with previous studies, failing the WMT was associated with being older and with higher self-reported level of depression (Erdodi, Abeare et al., 2018; Erdodi, Seke et al., 2017), but unrelated to sex, education, lateral dominance, and GCS. Patients who passed the WMT produced mean scores in the average range on ability tests, in line with the cumulative evidence that a full recovery of cognitive functioning is the normative outcome 3 months after a mild TBI (Rohling et al., 2011, McCrea et al., 2008).

In contrast, those who failed the WMT had mean scores on neuropsychological tests in the borderline range. These contrasts were associated with effect sizes (*d* .64–1.56) that would be considered large even for experimental malingering paradigms (Rapport, Farchione, Coleman, & Axelrod, 1998; Rogers, Sewell, Martin, & Vitacco, 2003; Suhr & Boyer, 1999). Moreover, failing the WMT was associated with increased within-group variability on a test of concept formation and a derivative index of executive deficits, a phenomenon considered to be an emergent sign of non-credible responding probably owing to the heterogeneity in malingering strategy across individuals (Cottingham, Victor, Boone, Ziegler, & Zeller, 2014; Erdodi, Kirsch, et al., 2014; Erdodi, Pelletier, & Roth, 2018). This neurocognitive profile was replicated with remarkable consistency using the VI-7 and EI-5 as criterion PVTs.

The WMT produced a high rate of agreement (74.7–84.1%) with other criterion PVTs, providing objective empirical evidence against claims that the failure rate on the WMT is artificially inflated by false-positive errors. Finally, the strongest evidence comes from a careful examination of the neurocognitive profile of patients who failed the WMT: 94.7% of them had independent psychometric evidence of non-credible responding. Therefore, they are considered to

be true positives, bringing the actual FPR down to 5.3%, which is considered low by most standards (Boone, 2013; Donders & Strong, 2011; Larrabee, 2003).

Inevitably, the study also has a number of limitations. First, it did not examine FPR in any diagnosis apart from mild TBI. However, limiting the investigation to mild TBI was necessary to ensure high internal validity, and to produce findings that would likely generalize to one of the most contentious category of patients in terms of PVT profiles. This clinical population was chosen partly because it lends itself well to PVT studies due its propensity to PVT failures. It is a condition which has been extensively studied, leading to the conclusion that mild TBI does not produce permanent measurable neuropsychological impairment in credible examinees (Rohling et al., 2011; McCrea et al., 2008).

Thus, failure on a PVT designed to be very easy cannot be attributed to genuine cognitive impairment. In contrast, FPRs in individuals with severe cognitive impairment are more difficult to estimate. Although empirically supported algorithms to aid the clinical interpretation of PVT failures in patients with more severe cognitive impairment have been developed (based on profile analysis of tests like the WMT, MSVT, and NV-MSVT) and their clinical utility continues to be replicated by independent investigators (Alverson, O'Rourke, & Soble, 2019; Elbaum, Golan, Lupu, Wagner, & Braw, 2019; Lupu, Elbaum, Wagner, & Braw, 2018; Shelley-Tremblay, Eyer, & Hill, 2019; Tomer, Lupu, Golan, Wagner, & Braw, 2018), a detailed discussion of this built-in safeguard against false-positive errors within the Green family of PVTs is beyond the scope of this paper.

One by-product of the high sensitivity of the WMT, as seen in multiple studies based on experimental malingerers, is that more people will fail the WMT than less sensitive PVTs. Although it is a natural tendency to ask whether the higher failure rate on the WMT versus other PVTs might (at least partly) be due to an increased FPR on the WMT, this may reflect a bias inherent in the assessor, not the test (Erdodi & Lichtenstein, 2017). The evidence from the current study does not support this interpretation. On the contrary, the results add to the impression from many studies summarized in the introduction that individuals with mild TBI who fail the WMT are unlikely to be false positives.

**Funding Information** This research was supported by a Collaborative Research Grant from the Faculty of Arts, Humanities, and Social Sciences at the University of Windsor.

## Compliance with Ethical Standards

Relevant ethical guidelines regulating research involving human participants were followed throughout the project. All data collection, storage, and processing were done in compliance with the Helsinki Declaration.

**Conflict of Interest** The first and last authors are employed by the University of Windsor. In addition to that, they provide forensic

consultation and medicolegal assessments, for which they receive financial compensation. Dr. Green is the author of the WMT, MSVT, and NV-MSVT and the owner of Green's Publishing. However, the data analysis was performed and interpreted by the other authors.

## References

- Abeare, C., Messa, I., Whitfield, C., Zuccato, B., Casey, J. & Erdodi, L. (2018). Performance validity in collegiate football athletes at baseline neurocognitive testing. *Journal of Head Trauma Rehabilitation*.
- Abeare, C. A., Messa, I., Zuccato, B. G., Merker, B., & Erdodi, L. A. (2018). Prevalence of invalid performance on baseline testing for sport-related concussion by age and validity indicator. Advanced online publication. *JAMA Neurology*, 75, 697–703. <https://doi.org/10.1001/jamaneurol.2018.0031>.
- Abeare, C., Sabelli, A., Taylor, B., Holcomb, M., Dumitrescu, C., Kirsch, N., & Erdodi, L. (2019). The importance of demographically adjusted cutoffs: age and education bias in raw score cutoffs within the Trail Making Test. Advance online publication. *Psychological Injury and Law*. <https://doi.org/10.1007/s12207-019-09353>.
- Allen, L. M., & Green, P. (1999). *Severe TBI sample performance on CARB and the WMT; supplement to the CARB '97 and Word Memory Test Manuals*. North Carolina: Cognisyst.
- Alverson, W. A., O'Rourke, J. J. F., & Soble, J. R. (2019). The Word Memory Test genuine memory impairment profile discriminates genuine memory impairment from invalid performance in a mixed clinical sample with cognitive impairment. *The Clinical Neuropsychologist*, 21(2), 209–231. <https://doi.org/10.1080/13825580601025932>
- An, K. Y., Kaploun, K., Erdodi, L. A., & Abeare, C. A. (2017). Performance validity in undergraduate research participants: a comparison of failure rates across tests and cutoffs. *The Clinical Neuropsychologist*, 31(1), 193–206. <https://doi.org/10.1080/13854046.2016.1217046>
- Ashendorf, L., Clark, E. L., & Sugarman, M. A. (2017). Performance validity and processing speed in a VA polytrauma sample. *The Clinical Neuropsychologist*, 31(5), 857–866. <https://doi.org/10.1080/13854046.2017.1285961>.
- Blaskewitz, N., Merten, T., & Brockhaus, R. (2009). Detection of suboptimal effort with the Rey Complex Figure Test and Recognition Trial. *Applied Neuropsychology*, 16, 54–61.
- Boone, K. B. (2013). *Clinical practice of forensic neuropsychology*. New York: Guilford.
- Bowden, S. C., Shores, E. A., & Mathias, J. L. (2006). Does effort suppress cognition after brain injury? A re-examination of the evidence for the Word Memory Test. *The Clinical Neuropsychologist*, 20, 858–872.
- Brockhaus, R., & Merten, T. (2004). Neuropsychologische diagnostik suboptimalen leistungsverhaltens mit dem Word Memory Test (Neuropsychological assessment of suboptimal performance: the Word Memory Test). *Nervenarzt*, 75, 882–887.
- Brockhaus, R., & Peker, O. (2003). Testing effort in Turkish-speaking subjects: validation of a translation of the Word Memory Test (WMT). *Journal of the International Neuropsychological Society*, 9, 503–590.
- Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury*, 22(12), 960–971.
- Carone, D. A. (2014). Young child with severe brain volume loss easily passes the Word Memory Test and Medical Symptom Validity Test: implication for mild TBI. *The Clinical Neuropsychologist*, 28, 146–162.
- Carone, D. A., Green, P., & Drane, D. L. (2013). Word Memory Test profiles in two cases with surgical removal of the left anterior hippocampus and parahippocampal gyrus. *Applied Neuropsychology: Adult*, 21(2), 155–160.

- Clark, A. L., Amick, M. M., Fortier, C., Millberg, W. P., & McGlinchey, R. E. (2014). Poor performance validity predicts clinical characteristics and cognitive test performance of OEF/OIF/OND veterans in a research setting. *The Clinical Neuropsychologist*, 28(5), 802–825.
- Conder, R., Allen, L., & Cox, D. (1992). *Computerized assessment of response bias test manual*. Durham: Cognisyst.
- Coleman, R. D., Rapport, L. J., Millis, S. R., Ricker, J. H., & Farchine, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning test. *Journal of Clinical and Experimental Neuropsychology*, 20(2), 201–210.
- Cottingham, M. E., Victor, T. L., Boone, K. B., Ziegler, E. A., & Zeller, M. (2018). Apparent effect of type of compensation seeking (disability vs. litigation) on performance validity test scores may be due to other factors. *The Clinical Neuropsychologist*, 28(6), 1030–1047. <https://doi.org/10.1080/13854046.2014.951397>.
- Curtis, K. L., Greve, K. W., Bianchini, K. J., & Brennan, A. (2006). California Verbal Learning Test indicators of malingered neurocognitive dysfunction: Sensitivity and specificity in traumatic brain injury. *Assessment*, 13(1), 46–61.
- DeFlippis, N. A., & McCampbell, E. (1979, 1991, 1997). Manual for the Booklet Category Test. Odessa: *Psychological Assessment Resources*.
- Delis, D. C., & Wetter, S. R. (2007). Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology*, 22, 589–604.
- Demakis, G. J., Gervais, R. O., & Rohling, M. L. (2008). The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *The Clinical Neuropsychologist*, 22(5), 879–895.
- DiCarlo, M. A., Gfeller, J. D., & Oliveri, M. V. (2000). Effects of coaching on detecting feigned cognitive impairment with the Category Test. *Archives of Clinical Neuropsychology*, 15(5), 399–413.
- Donders, J., & Strong, C. H. (2011). Embedded effort indicators on the California Verbal Learning Test – second edition (CVLT-II): an attempted cross-validation. *The Clinical Neuropsychologist*, 25, 173–184.
- Elbaum, T. Golan, L., Lupu, T., Wagner, M., & Braw, Y. (2019). Establishing supplementary response time validity indicators in the Word Memory Test (WMT) and directions for future research. Advanced online publication. *Applied Neuropsychology: Adult*. <https://doi.org/10.1080/23279095.2018.1555161>.
- Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B. T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017). WAIS-IV processing speed scores as measures of non-credible responding – the third generation of embedded performance validity indicators. *Psychological Assessment*, 29(2), 148–157. <https://doi.org/10.1037/pas0000319>.
- Erdodi, L. A., Abeare, C. A., Medoff, B., Seke, K. R., Sagar, S., & Kirsch, N. L. (2018). A single error is one too many: The Forced Choice Recognition trial on the CVLT-II as a measure of performance validity in adults with TBI. *Archives of Clinical Neuropsychology*, 33(7), 845–860. <https://doi.org/10.1093/arclin/acx110>.
- Erdodi, L. A., Dunn, A. G., Seke, K. R., Charron, C., McDermott, A., Enache, A., Maytham, C., & Hurtubise, J. (2018). The Boston Naming Test as a measure of performance validity. *Psychological Injury and Law*, 11, 1–8. <https://doi.org/10.1007/s12207-017-9309-3>
- Erdodi, L. A., Hurtubise, J. L., Charron, C., Dunn, A., Enache, A., McDermott, A., & Hirst, R. (2018). The D-KEFS Trails as performance validity tests. *Psychological Assessment*, 30(8), 1081–1095.
- Erdodi, L. A., Kirsch, N. L., Lajiness-O'Neill, R., Vingilis, E., & Medoff, B. (2014). Comparing the Recognition Memory Test and the Word Choice Test in a mixed clinical sample: Are they equivalent? *Psychological Injury and Law*, 7(3), 255–263. <https://doi.org/10.1007/s12207-014-9197-8>.
- Erdodi, L. A., Kirsch, N. L., Sabelli, A. G., & Abeare, C. A. (2018). The grooved pegboard test as a validity indicator – a study on psychogenic interference as a confound in performance validity research. *Psychological Injury and Law*, 11(4), 307–324. <https://doi.org/10.1007/s12207-018-9337-7>.
- Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: an emerging paradox of embedded validity indicators. *The Clinical Neuropsychologist*, 31(6–7), 1029–1046. <https://doi.org/10.1080/13854046.2017.1323119>.
- Erdodi, L. A., & Lichtenstein, J. D. (2019). Information processing speed tests as PVTs. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment. A neuropsychological perspective*. New York: Guilford.
- Erdodi, L. A., Lichtenstein, J. D., Rai, J. K., & Flaro, L. (2017). Embedded validity indicators in Conners' CPT-II: do adult cutoffs work the same way in children? *Applied Neuropsychology: Child*, 6(4), 335–363. <https://doi.org/10.1080/21622965.2016.1198908>.
- Erdodi, L. A., Nussbaum, S., Sagar, S., Abeare, C. A., & Schwartz, E. S. (2017). Limited English proficiency increases failure rates on performance validity tests with high verbal mediation. *Psychological Injury and Law*, 10(1), 96–103.
- Erdodi, L. A., Pelletier, C. L., & Roth, R. M. (2018). Elevations on select Conners' CPT-II scales indicate noncredible responding in adults with traumatic brain injury. *Applied Neuropsychology: Adult*, 25(1), 19–28. <https://doi.org/10.1080/23279095.2016.1232262>
- 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., & Roth, R. M. (2017). Low scores on BDAE Complex Ideational Material are associated with invalid performance in adults without aphasia. *Applied Neuropsychology: Adult*, 24(3), 264–274. <https://doi.org/10.1080/23279095.2017.1298600>.
- Erdodi, L. A., Seke, K. R., Shahein, A., Tyson, B. T., Sagar, S., & Roth, R. M. (2017). Low scores on the Grooved Pegboard Test are associated with invalid responding and psychiatric symptoms. *Psychology and Neuroscience*, 10(3), 325–344. <https://doi.org/10.1037/pnc0000103>.
- Erdodi, L. A., Sagar, S., Seke, K., Zuccato, B. G., Schwartz, E. S., & Roth, R. M. (2018a). The Stroop Test as a measure of performance validity in adults clinically referred for neuropsychological assessment. *Psychological Assessment*, 30(6), 755–766. <https://doi.org/10.1037/pas0000525>.
- Erdodi, L. A., Tyson, B. T., Abeare, C. A., Zuccato, B. G., Rai, J. K., Seke, K. R., Sagar, S., & Roth, R. M. (2018b). Utility of critical items within the Recognition Memory Test and Word Choice Test. *Applied Neuropsychology: Adult*, 25(4), 327–339. <https://doi.org/10.1080/23279095.2017.1298600>.
- Erdodi, L. A., Tyson, B. T., Shahein, A., Lichtenstein, J. D., Abeare, C. A., Pelletier, C. L., Zuccato, B. G., Kucharski, B., & Roth, R. M. (2017b). The power of timing: adding a time-to-completion cutoff to the Word Choice Test and Recognition Memory Test improves classification accuracy. *Journal of Clinical and Experimental Neuropsychology*, 39(4), 369–383. <https://doi.org/10.1080/13803395.2016.1230181>.
- Etherton, J. L., Bianchini, K. J., Heinly, M. T., & Greve, K. W. (2006). Pain, malingering, and performance on the WAIS-III Processing Speed Index. *Journal of Clinical and Experimental Neuropsychology*, 28(7), 1218–1237.
- Fritze, E. (2003). Validierung des Word Memory Test in Türkische sprache (Validation of the WMT in Turkish speakers). *External examiner, Brockhaus, R., Diplomarbeit* (Thesis), Ruhr-Universität Bochum, Dortmund, Germany.
- Gervais, R. O., Russell, A. S., Green, P., Allen, L. M., Ferrari, R., & Pieschl, S. D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology*, 28, 1892–1899.

- Gligorović, M., & Buha, N. (2013). Conceptual abilities of children with mild intellectual disability: analysis of Wisconsin Card Sorting Test performance. *Journal of Intellectual and Developmental Disabilities, 38*(2), 134–140. <https://doi.org/10.3109/13668250.2013.772956>.
- Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2009). Word Memory Test performance in amnesic patients with hippocampal damage. *Neuropsychology, 23*(4), 529–534.
- Green, P. (2003). *Green's Word Memory Test*. Edmonton, Canada: Green's Publishing.
- Green, P. (2004). *Green's medical symptom validity test (MSVT) for Microsoft windows user's manual*. Kelowna: Green's Publishing.
- Green, P. (2007). Spoiled for choice: Making comparisons between forced-choice effort tests. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment* (pp. 50–77). New York: Guilford Press.
- Green, P., & Allen, L. (1999). *Performance of neurological patients on the Word Memory Test (WMT) and computerized assessment of response bias (CARB); supplement to the Word Memory Test and CARB '97 Manuals*. North Carolina: Cognisyst.
- Green, P., & Astner, K. (1995). *Manual for the oral word memory test*. North Carolina: Cognisyst.
- Green, P., & Flaro, F. (2015). Results from three performance validity tests (PVTs) in adults with intellectual deficits. *Applied Neuropsychology: Adult, 22*(4), 293–303.
- Green, P., & Flaro, L. (2016). Results from three performance validity tests in children with intellectual disability. *Applied Neuropsychology: Child, 5*, 25–34.
- Green, P., Flaro, L., Brockhaus, R., & Montijo, J. (2012). Performance on the WMT, MSVT, and NV-MSVT in children with developmental disabilities and in adults with mild traumatic brain injury. In A. Horton, & Reynolds, C. (Eds.) New York, USA: Springer.
- Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury, 23*(9), 741–750.
- Green, P., Lees-Haley, P., & Allen, L. M. (2003). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology, 2*, 97–124.
- Green, P., & Merten, T. (2013). Noncredible explanations of noncredible symptom validity test failure. In D. Carone & S. Bush (Eds.), *Mild traumatic brain injury: symptom validity assessment and malingering*. New York: Springer Publishing.
- Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology, 18*(2), 86–94.
- Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*(12), 1045–1060.
- Greiffenstein, M. F., & Baker, W. J. (2008). Validity testing in dually diagnosed posttraumatic stress disorder and mild closed head injury. *The Clinical Neuropsychologist, 22*, 565–582.
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingering amnesia measures with a large clinical sample. *Psychological Assessment, 6*(3), 218–224.
- Greve, K., Ord, J., Curtis, K., Bianchini, K., & Brennan, B. (2008). Detecting malingering in traumatic brain injury and chronic pain: a comparison of three forced choice symptom validity tests. *The Clinical Neuropsychologist, 22*, 896–918.
- Greve, K. W., Heinly, M. T., Bianchini, K. J., & Love, J. M. (2009). Malingering detection with the Wisconsin Card Sorting Test in mild traumatic brain injury. *The Clinical Neuropsychologist, 23*, 343–362.
- Grimes, D. A., & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *The Lancet, 365*(9469), 1500–1505.
- Hall, V., Worthington, A., & Venables, K. (2014). A UK pilot study: the specificity of the Word Memory Test effort sub-tests in acute minimal to mild head injury. *Journal of Neuropsychology, 8*(2), 216–230.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). Revised comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults. Lutz: Psychological Assessment Resources.
- Henry, G. K., Helbronner, R. L., Suhr, J., Gornbein, J., Wagner, E., & Drane, D. L. (2018). Illness perceptions predict cognitive performance validity. *Journal of the International Neuropsychological Society, 24*, 1–11.
- Hill, A. B. (1965). The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine, 58*, 295–300.
- Hurtubise, J. L., Scavone, A., Sagar, S., & Erdodi, L. A. (2017). Psychometric markers of genuine and feigned neurodevelopmental disorders in the context of applying for academic accommodations. *Psychological Injury and Law, 10*, 121–137.
- Jones, A. (2013). Test of memory malingering: cutoff scores for psychometrically defined malingering groups in a military sample. *The Clinical Neuropsychologist, 27*(6), 1043–1059.
- King, J. H., Sweet, J. J., Sherer, M., Curtiss, G., & Vanderploeg, R. D. (2002). Validity indicators within the Wisconsin Card Sorting Test: Application of new and previously researched multivariate procedures in multiple traumatic brain injury samples. *The Clinical Neuropsychologist, 16*(4), 506–523.
- Kulas, J. F., Axelrod, B. N., & Rinaldi, A. R. (2014). Cross-validation of supplemental Test of Memory Malingering Scores as performance validity measures. *Psychological Injury and Law, 7*(3), 236–244.
- Larochette, A. C., & Harrison, A. G. (2012). Word Memory Test performance in Canadian adolescents with learning disabilities: a preliminary study. *Applied Neuropsychology: Child, 1*(1), 38–47.
- 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. (2012). Assessment of malingering. In G. J. Larrabee (Ed.), *Forensic neuropsychology: a scientific approach* (Second ed., pp. 116–159). New York: Oxford University Press.
- Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology, 29*, 364–373.
- Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: reply to Russell. *Clinical Neuropsychology, 23*(5), 841–849.
- Lichtenstein, J. D., Flaro, L., Baldwin, F. S., Rai, J., & Erdodi, L. A. (2019). Further evidence for embedded performance validity tests in children within the Conners' Continuous Performance Test - second edition. *Developmental Neuropsychology, 44*(2), 159–171. <https://doi.org/10.1080/87565641.2019.1565535>.
- Lichtenstein, J. D., Holcomb, M., & Erdodi, L. A. (2018). One-minute PVT: further evidence for the utility of the California Verbal Learning Test—Children's Version Forced Choice Recognition Trial. *Journal of Pediatric Neuropsychology, 4*, 94–104.
- Lupu, T., Elbaum, T., Wagner, M., & Braw, Y. (2018). Enhanced detection of feigned cognitive impairment using per item response time measurements in the Word Memory Test. *Applied Neuropsychology: Adult, 25*(6), 532–542. <https://doi.org/10.1080/23279095.2017.1341410>.
- Maccocchi, S. N., Seel, R. T., Yi, A., & Small, S. (2017). Medical symptom validity test performance following moderate-severe traumatic brain injury: expectations based on orientation log classification. *Archives of Clinical Neuropsychology, 32*(3), 339–348.
- Mathias, C. W., Greve, K. W., Bianchini, K. J., Houston, R. J., & Crouch, J. A. (2002). Detecting malingered neurocognitive dysfunction

- using the reliable digit span in traumatic brain injury. *Assessment*, 9(3), 301–308.
- McCrea, M. A. (2008). *Mild traumatic brain injury and post-concussion syndrome: the new evidence*. Oxford University Press.
- Meyers, J., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology*, 26(1), 8–15.
- Millis, S. R., Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment*, 7(4), 463–471.
- 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.
- Pearson. (2009). *Advanced Clinical Solutions for the WAIS-IV and WMS-IV – Technical Manual*. San Antonio: Author.
- Persinger, V. C., Whiteside, D. M., Bobova, L., Saigal, S. D., Vannucci, M. J., & Basso, M. R. (2018). Using the California Verbal Learning Test, Second Edition as an embedded performance validity measure among individuals with TBI and individuals with psychiatric disorders. *The Clinical Neuropsychologist*, 32(6), 1039–1053. <https://doi.org/10.1080/13854046.2017.1419507>.
- Rapport, L. J., Farchione, T. J., Coleman, R. D., & Axelrod, B. N. (1998). Effects of coaching on malingered motor function profiles. *Journal of Clinical and Experimental Neuropsychology*, 20(1), 89–97.
- Rabin, L. A., Spadaccini, A. T., Brodale, D. L., Grant, K. S., Elbulok, M. M., & Barr, W. B. (2014). Utilization rates of tests & test batteries among clinical psychologists in the United States & Canada. *Professional Psychology Research & Practice*, 45(5), 368–377.
- Rai, J., An, K. Y., Charles, J., Ali, S., & Erdodi, L. A. (2019). Introducing a forced choicerecognition trial to the Rey Complex Figure Test. Advance online publication. *Psychology and Neuroscience*. <https://doi.org/10.1037/pne0000175>.
- Rai, J., & Erdodi, L. (2019). The impact of criterion measures on the classification accuracy of TOMM-1. Advance online publication. *Applied Neuropsychology: Adult*. <https://doi.org/10.1080/23279095.2019.161.1613994>.
- Reese, C. S., Suhr, J. A., & Riddle, T. L. (2012). Exploration of malingering indices in the Wechsler Adult Intelligence Scale – Fourth Edition Digit Span subtest. *Archives of Clinical Neuropsychology*, 27, 176–181.
- Richman, J., Green, P., Gervais, R., Flaro, L., Merten, T., Brockhaus, R., & Ranks, D. (2006). Objective tests of symptom exaggeration in independent medical evaluations. *Journal of Occupational and Environmental Medicine*, 48(3), 303–311.
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, 10, 160–177.
- Rohling, M., Binder, L., Demakis, G., Larrabee, G., Ploetz, L., & Langhinrichsen-Rohling, J. (2011). A meta-analysis of neuropsychological outcome after mild traumatic brain injury: re-analyses and reconsiderations of Binder et al., Frencham et al., and Pertab et al. *The Clinical Neuropsychologist*, 25(4), 608–623.
- Rohling, M. L., & Demakis, G. J. (2010). Bowden, Shores, & Mathias (2006): failure to replicate or just failure to notice. Does effort still account for more variance in neuropsychological test scores than TBI severity? *The Clinical Neuropsychologist*, 24(1), 119–136.
- Rohling, M. L., Green, P., Allen, L., & Iverson, G. L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Archives of Clinical Neuropsychology*, 17(3), 205–222.
- Shelley-Tremblay, J. F., Eyer, J. C., & Hill, B. D. (2019). A laboratory Word Memory Test analogue differentiates intentional feigning from true responding using the P300 event-related potential. *Brain Science*, 9(5), 109. <https://doi.org/10.3390/brainsci9050109>.
- Shura, R. D., Miskey, H. M., Rowland, J. A., Yoash-Gatz, R. E., & Denning, J. H. (2016). Embedded performance validity measures with postdeployment veterans: Crossvalidation and efficiency with multiple measures. *Applied Neuropsychology: Adult*, 23, 94–104. <https://doi.org/10.1080/23279095.2015.1014556>
- 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.
- Slick, D. J., Iverson, G. L., & Green, P. (2000). California Verbal Learning Test indicators of suboptimal performance in a sample of head-injury litigants. *Journal of Clinical and Experimental Neuropsychology*, 22(4), 569–579.
- Suhr, J. A. (2003). Neuropsychological impairment in fibromyalgia. Relation to depression, fatigue, and pain. *Journal of Psychosomatic Research*, 55, 321–329.
- Suhr, J. A., & Boyer, D. (1999). Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. *Journal of Clinical and Experimental Psychology*, 21(5), 701–708. <https://doi.org/10.1076/jcen.21.5.701.868>
- Sweet, J. J., Wolfe, P., Sattlberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S., & Nies, K. J. (2000). Further investigation of traumatic brain injury versus insufficient effort with the California Verbal Learning Test. *Archives of Clinical Neuropsychology*, 15(2), 105–113.
- Tenhula, W. N., & Sweet, J. J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *The Clinical Neuropsychologist*, 10(1), 104–116.
- Tombaugh, T. N. (1996). *Test of Memory Malingering*. New York: Multi-Health Systems.
- Tomer, E., Lupu, T., Golan, L., Wagner, M., & Braw, Y. (2018). Eye tracking as a mean to detect feigned cognitive impairment in the Word Memory Test. Advance online publication. *Applied Neuropsychology: Adult*. <https://doi.org/10.1080/23279095.2018.1480483>.
- Trueblood, W. (1994). Qualitative and quantitative characteristics of malingering and other invalid WAIS-R and clinical memory data. *Journal of Clinical and Experimental Neuropsychology*, 14(4), 697–607. <https://doi.org/10.1080/01688639408402671>.
- Tydecks, S., Merten, T., & Gubbay, J. (2006). The Word Memory Test and the One-in-Five-Test in an analogue study with Russian speaking participants. *International Journal of Forensic Psychology*, 1(3), 29–37.
- Tyson, B. T., Baker, S., Greenacre, M., Kent, K. J., Lichtenstein, J. D., Sabelli, A., & Erdodi, L. A. (2018). Differentiating epilepsy from psychogenic nonepileptic seizures using neuropsychological test data. *Epilepsy & Behavior*, 87, 39–45.
- Warrington, E. K. (1984). *Recognition Memory Test: RMT (Words) [Test Booklet]*. NFER-Nelson Publishing Company.
- Whiteside, D., Wald, D., & Busse, M. (2011). Classification accuracy of multiple visual spatial measures in the detection of suspect effort. *The Clinical Neuropsychologist*, 25, 287–301.
- Williamson, D. J., Holsman, M., Chaytor, N., Miller, J. W., & Drane, D. L. (2012). Abuse, not financial incentive, predicts non-credible cognitive performance in patients with psychogenic non-epileptic seizures. *The Clinical Neuropsychologist*, 26(4), 588–598.
- Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J., & Sweet, J. J. (2010). Effort indicators within the California Verbal Learning Test-II (CVLT-II). *The Clinical Neuropsychologist*, 24(1), 153–168.