Check for updates

Demographically Adjusted Validity Cutoffs on the Finger Tapping Test Are Superior to Raw Score Cutoffs in Adults with TBI

Laszlo A Erdodi [1] · Brian Taylor [2] · Alana G Sabelli [2] · Malayna Malleck [1] · Ned L Kirsch [3] · Christopher A Abeare [1]

Abstract

This study was designed to develop validity cutoffs within the Finger Tapping Test (FTT) using demographically adjusted T-scores, and to compare their classification accuracy to existing cutoffs based on raw scores. Given that FTT performance is known to vary with age, sex, and level of education, failure to correct for these demographic variables poses the risk of elevated false positive rates in examinees who, at the level of raw scores, have inherently lower FTT performance (women, older, and less educated individuals). Data were collected from an archival sample of 100 adult outpatients (M_{Age} = 38.8 years, $M_{Education}$ = 13.7 years, 56% men) consecutively referred for neuropsychological assessment at an academic medical center in the Midwestern USA after sustaining a traumatic brain injury (TBI). Performance validity was psychometrically defined using the Word Memory Test and two validity composites based on five embedded performance validity indicators. Previously published raw score-based validity cutoffs disproportionately sacrificed sensitivity (.13–.33) for specificity (.98–1.00). Worse yet, they were confounded by sex and education. Newly introduced demographically adjusted cutoffs ($T \leq 33$ for the dominant hand, $T \leq 37$ for both hands) produced high levels of specificity (.89–.98) and acceptable sensitivity (.36–.55) across criterion measures. Equally importantly, they were robust to injury severity and demographic variables. The present findings provide empirical support for a growing trend of demographically adjusted performance validity cutoffs. They provide a practical and epistemologically superior alternative to raw score cutoffs, while also reducing the potential bias against examinees inherently vulnerable to lower raw score level FTT performance.

Keywords Finger tapping test · Demographically adjusted cutoffs · Performance validity · Erdodi Index · Modality specificity

Valid interpretation of neuropsychological test results relies on the assumption that examinees perform to the best of their abilities. As such, it has become a broadly accepted standard of practice to routinely assess the credibility of examinees' responding (Bush et al., 2005; Heilbronner et al., 2009). Performance validity may also fluctuate over the course of a long testing session as a function of idiosyncratic expression of cognitive deficits and/or malingering strategy (Cottingham,

Victor, Boone, Ziegler, & Zeller, 2014; Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014a). Detection methods must be robust enough to identify a broad range of abnormal test taking behaviors that may indicate non-credible responding.

The most effective strategy for a comprehensive assessment of performance validity is using multivariate models (An, Kaploun, Erdodi, & Abeare, 2017; Boone, 2013; Erdodi et al., 2018c; Larrabee, 2008; Tyson et al., 2018). Therefore, it is important to administer several different measures of performance validity dispersed throughout an assessment (Boone, 2009; Bush, Heilbronner, & Ruff, 2014; Chafetz et al., 2015; Schutte, Axelrod, & Montoya, 2015). Performance validity may be compromised by a wide range of factors, such as fatigue (Kalfon, Gal, Shorer, & Ablin, 2016; Suhr, 2003), lack of interest (An et al., 2017, 2019; Rai, An, Charles, Ali, & Erdodi, 2019), cogniphobia (Suhr & Spickard, 2012), illness perception (Henry et al., 2018), acute emotional distress Clark, Amick, Fortier, Millberg, & McGlinchey, 2014; Erdodi et al., 2016), failing to appreciate

✉ Laszlo A Erdodi
    lerdodi@gmail.com

[1] Department of Psychology, Neuropsychology Track, University of Windsor, 168 Chrysler Hall South, 401 Sunset Ave, Windsor, ON N9B 3P4, Canada

[2] Behaviour-Cognition-Neuroscience Program, University of Windsor, Windsor, ON, Canada

[3] Department of Physical Medicine and Rehabilitation, Adult Outpatient Neurehabilitation Program, University of Michigan, Ann Arbor, MI, USA

✷ Springer

the importance of demonstrating highest level of neurocognitive ability (Abeare, Messa, Zuccato, Merker, & Erdodi, 2018b; Abeare et al., 2018a), complex trauma history (Costabile, Bilo, DeRosa, Pane, & Sacca, 2018; Williamson, Holsman, Chaytor, Miller, & Drane, 2012), deficits in adaptive functioning (Lippa, Agbayani, Hawes, Jokic, & Caroselli, 2014), or outright malingering (Boone, 2013; Larrabee, 2012). Regardless of the specific cause, non-credible responding is a significant threat to the validity of test data that can undermine the validity of the clinical interpretation.

Clinicians' judgment about credibility of performance has long been known to be inaccurate (Heaton, Smith, Lehman, & Vogt, 1978), despite (or perhaps because) high level of confidence in their ability to detect non-credible performance (Guilmette, 2013). Therefore, objective empirical methods are needed to evaluate performance validity (Chafetz et al., 2015; Larrabee, 2014). By design, there are two main categories of instruments developed to monitor performance validity: free-standing performance validity tests (PVTs) and embedded validity indicators (EVIs).

Free-standing PVTs were created with the specific and sole purpose of monitoring the credibility of test taking behavior. Consequently, they provide little (if any) information on cognitive ability. They are usually simple tasks that were designed to appear more difficult than they actually are. In addition to low cognitive demand, the threshold of failure is often established so that the majority of examinees with severe genuine impairment are able to pass it (Allen & Green, 1999; Greve, Bianchini, & Doane, 2006; Green & Flaro, 2003). Given the combination of an objectively easy task and typically highly conservative cutoffs (Erdodi & Rai, 2017; Greve, Etherton, Ord, Bianchini, & Curtis, 2009; Jones, 2013; Kulas, Axelrod, & Rinaldi, 2014), failing free-standing PVTs provides strong evidence of non-credible responding that is unlikely to be attributable to genuine neurological impairment. However, the practice of deeming an entire neurocognitive profile *invalid* based on failures on free-standing PVTs represents an inferential leap that has been criticized on both logical (An et al., 2019; Bigler, 2012) and empirical grounds (Erdodi, Pelletier, & Roth, 2018e; Erdodi & Roth, 2017; Locke, Smigielski, Powell, & Stevens, 2008).

In contrast, EVIs are derived from traditional neuropsychological tests subsequently co-opted as PVTs (Committee on Psychological Testing, 2015). EVIs have several advantages over free-standing PVIs, including being cost-effective (Boone, 2013; Erdodi, Abeare, Lichtenstein, Erdodi, & Linnea, 2017; Erdodi, Hurtubise, et al., 2018c) and more resistant to coaching, in that examinees are less likely to be able to identify them as PVTs in comparison to their free-standing counterparts (Chafetz et al., 2015; Miele, Gunner, Lynch, & McCaffrey, 2012). In addition, relying on EVIs reduces the demand on the examinee's mental stamina, which can be particularly useful when testing young children and medically or emotionally fragile patients (Lichtenstein et al., 2017). The main liability inherent in relying on EVIs to determine the credibility of a given response set is that separating ability from performance validity is more difficult and not without controversy (Erdodi & Lichtenstein, 2017).

The Finger Tapping Test (FTT; Reitan, 1969) is a measure of simple motor speed that has subsequently demonstrated its potential for use as an EVI (Heaton et al., 1978; Mittenberg, Rotholc, Russell, & Heilbronner, 1996; Rapport, Farchione, Coleman, & Axelrod, 1998). During the FTT, examinees are instructed to tap a lever attached to a mechanical counter using their extended index finger for 10 s at a time as fast as they can, over at least five consecutive trials. This procedure is then repeated for the non-dominant hand. Once the FTT is completed, the trials are averaged to create a summary score for each hand, which is used to establish the individual's relative standing. The raw score performance of the FTT is used to compare oscillation speed between the dominant and non-dominant hands in order to infer the integrity of the cortical motor areas and the efferent motor pathways (Schatz, 2011).

Historically, the potential of motor measures to evaluate credible performance has received less attention than measures assessing other cognitive domains, such as memory—especially the two-alternative forced choice recognition paradigm (Baker, Connery, Kirk, & Kirkwood, 2014; Bauer, Yantz, Ryan, Warden, & McCaffrey, 2005; Blaskewitz, Merten, & Brockhaus, 2009; Bortnik et al., 2010; Greve et al., 2009; Lu, Boone, Cozolino, & Mitchell, 2003; Reedy et al., 2013; Wolfe et al., 2010). However, their ability to separate valid from invalid response sets has longstanding empirical support. The FTT demonstrated its ability to perform as an EVI in multiple studies in which participants feigned head injuries to determine if individuals with non-credible performance can be differentiated from those with true head injuries (Greiffenstein, Baker, & Gola, 1996; Heaton et al., 1978; Rapport et al., 1998). Feigning participants demonstrated abnormally poor and inconsistent performance patterns (i.e., performing worse on simple motor tests, such as the FTT, than on more complex measures, such as the Grooved Pegboard Test; Matthews & Klove, 1964). Overall, the existing evidence suggests that the FTT can help separate examinees who are feigning injuries from those with genuine motor deficits. However, early studies stopped short of introducing specific cutoffs to separate credible from non-credible response sets (Erdodi et al., 2017c).

Eventually, normative data that included demographic corrections (i.e., taking into account differences in performance due to age, sex, and education) were published as an alternative to the raw score cutoffs typically used to determine impairment (Heaton, Grant, & Matthews, 1991). Naturally, performance on the FTT is influenced by innate motor skills and, by extension, demographic variables. Specifically, previous research has shown that older age was associated with

decreased finger tapping speed (Shimoyama et al., 2012). Similarly, at group level, males typically outperformed females (Arnold et al., 2005; Ruff & Parker, 1993). Also, examinees with more formal education demonstrated better performance on the FTT (Homann et al., 2003). Such examinee characteristics with known influence on FTT performance can inadvertently bias validity cutoffs by artificially inflating failure rates in specific demographic groups that are at inherently higher risk for low finger tapping scores (i.e., examinees who are women, older, and have less education).

The FTT's integrity as a performance validity indicator hinges on its potential to differentiate between ability, which appears to vary with demographic factors, and performance validity, which is thought to be independent of demographic factors (Boone, 2013; Lichtenstein, Erdodi, Rai, Mazur-Mosiewicz, & Flaro, 2018a; Lichtenstein, Holcomb, & Erdodi, 2018b). The credibility of an individual response set is determined based on binary validity cutoffs (i.e., *Pass/Fail*). Although existing FTT validity cutoffs have typically been calibrated to clear the minimum acceptable threshold for specificity (.84; Larrabee, 2003; .90; Boone, 2013), they have historically been based exclusively on raw scores.

Larrabee (2003) published validity cutoffs on the FTT that combine dominant and non-dominant hand raw scores. However, these cutoffs do not account for the effect of demographic variables. The only demographic factor currently used to adjust cutoffs for the FTT is sex: Arnold et al. (2005) were the first to formally recognize the need for a separate set of cutoffs for men (higher) and women (lower) to maintain comparable classification accuracy. In more recent research, Axelrod, Meyers, and Davis (2014) explored alternative cutoff scores but did not adjust for any demographic variables.

Previously published cutoffs that exclusively rely on raw scores may inadvertently introduce systematic error in the interpretation of test results. Namely, groups that demonstrate naturally lower performance on this task (i.e., women, older examinees, less educated individuals) are at higher risk for false positive error. The present study was designed to take the next logical step in developing the FTT as an EVI by offering more comprehensive demographically adjusted validity cutoffs. Therefore, FTT raw scores were transformed to T-scores corrected for race, age, gender, and education using normative data published by Heaton, Miller, Taylor, and Grant (2004). Existing cutoffs based on raw scores and newly developed cutoffs based on T-scores were compared directly to evaluate their classification accuracy (i.e., their ability to separate valid and invalid responses). It was hypothesized that using demographically adjusted validity cutoffs for the FTT would achieve specificity and sensitivity values comparable to raw score cutoffs. In addition, we predicted that demographically adjusted FTT validity cutoffs would protect demographically disadvantaged examinees against false positive errors.

Method

Participants

The sample consisted of a consecutive case sequence of 100 patients evaluated following a TBI at an academic medical center in the Midwestern US (fifth author's institutional affiliation). All patients were clinically referred for a neuropsychological evaluation by their treating physician. Data on external incentives to appear impaired (Criterion A of Malingered Neurocognitive Dysfunction; Slick, Sherman, & Iverson, 1999) were not available for the majority of the sample. Instead, the credibility of neurocognitive profiles was established using psychometric methods. Patients with missing index fingers, acute orthopedic injury to their hand, or neurological conditions known to cause significant impairment in dexterity (hemiparesis, multiple sclerosis, Parkinson's disease) were excluded from the study to control for a major confound in evaluating the FTT's potential to serve as a PVT.

The majority of the participants were men (56%) and right-handed (91%). Mean age was 38.8 (SD = 14.9, range 17–74). Mean level of education was 13.7 years (SD = 2.6, range 7–20). Mean FSIQ was 92.7 (SD = 15.7, range 61–130). Although seven of the patients produced FSIQ < 70, none of them had a documented history of longstanding adaptive deficits required for a diagnosis of intellectual disability. In fact, two of them had a university diploma. All of them had psychometric evidence of normal or even above average cognitive functioning. Therefore, FSIQ < 70 in this subset of patients was likely a manifestation of non-credible responding.

As expected, the majority of head injuries were categorized as mild TBI (76%) by the assessing neuropsychologist according to commonly used guidelines (GCS > 13, LOC < 30 min, and PTA < 24 h; American Congress of Rehabilitation Medicine, 1993). Data on injury parameters were collected through clinical interview and the review of the medical records. The present sample overlaps with previous publications from the same research group designed to investigate different topics (Erdodi, Kirsch et al., 2018d; Erdodi, Abeare, Medoff et al., 2018; Erdodi, Roth et al., 2014b). None of these papers examined the FTT as a PVT.

Materials

Tests Administered

A core battery of neuropsychological tests was administered to all patients, including the Wechsler Adult Intelligence (WAIS-IV; Wechsler, 2008) and Memory Scales (WMS-IV; Wechsler, 2009), the California Verbal Learning Test – Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000), Conners' Continuous Performance Test – Second

Edition (CPT-II; Conners, 2004), the Trail Making Test (TMT A & B; Reitan, 1955; Reitan, 1958), verbal fluency (FAS & animals; Gladsjo et al., 1999; Newcombe, 1969), the Wisconsin Card Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curtis, 1993), the Grooved Pegboard Test (GPB; Matthews & Klove, 1964), and the Tactual Performance Test (TPT; Halstead, 1947). Premorbid functioning was estimated using the single word reading subtest on the Wide Range Achievement Test (WRAT-4; Wilkinson & Robertson, 2006) and the Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4; Dunn & Dunn, 2007). Manual dexterity was measured with the FTT and GPB, using demographically adjusted T-scores based on norms by Heaton et al. (2004).

Administration protocols for the FTT vary (Strauss, Sherman, & Spreen, 2006). Patients in the present sample were administered the FTT following the instructions by Reitan and Wolfson (1985): five consecutive trials with the dominant hand, followed by five consecutive trials with the non-dominant hand. If the range of raw scores across the five trials was within 5 points, the test was discontinued for that hand. If performance was more variable, further trials were administered until the "five (trials) within five (range)" criterion was met, or 10 trials were reached.

Criterion PVTs

The main free-standing PVT was the computer-administered version of the Word Memory Test (WMT; Green, 2003). The Genuine Memory Impairment Profile (GMIP) was not applied as an exclusionary clause after failing the WMT, for a number of reasons. First, although earlier investigations demonstrated that careful screening for both non-credible responding and genuine memory deficits is needed to detect a dose-response relationship between TBI severity and memory skills (Donders & Strong, 2013), the present study was explicitly focused on performance validity and non-memory-based EVIs. Second, previous research found that a quarter of graduate students asked to feign dementia qualified for the GMIP (Armistead-Jehle & Denney, 2015), suggesting that it is not specific to genuine impairment. Third, injury severity as a potential confound for PVT failure was explicitly investigated in this sample.

To meet the broadly accepted practice guideline of continuously monitoring performance validity throughout the assessment (Boone, 2009; Bush et al., 2014; Chafetz et al., 2015; Schutte et al., 2015), two validity composites were created ("Erdodi Index Five" or EI-5), following the methodology outlined by Erdodi (2019). The first one was based on PVTs embedded within measures of psychomotor processing speed and was labeled "EI-5$_{PSP}$" The second validity composite was based on PVTs embedded within measures of attention, working

memory, and auditory verbal memory, and was labeled "EI-5$_{AWM}$."

Each EI-5 component was recoded onto a 4-point ordinal scale, where 0 was defined as an incontrovertible *Pass*, and 3 as an incontrovertible *Fail*. Psychometrically, an EI-5 value of 1 is anchored to failing the most liberal (i.e., high sensitivity, low specificity) cutoff, which is typically associated with a base rate of failure (BR$_{Fail}$) around 25% (Pearson, 2009). The cutoffs for EI-5 values 2 and 3 are either determined on rational grounds in combination with the existing research literature, or in the case of scales with wide range, adjusted to correspond to BR$_{Fail}$ of 10 and 5, respectively (Table 1; Erdodi, Kirsch, Sabelli, & Abeare, 2018d). In other words, as the EI-5 value increases, so does the confidence in correctly classifying a given response set as *invalid*. By design, the EI-5 captures both the *number* and *extent* of PVT failures (Erdodi et al., 2018b).

The value of the EI-5 composite is obtained by summing its recoded components. As such, it can range from 0 (i.e., all five constituent PVTs were passed at the most liberal cutoff) to 15 (i.e., all five constituent PVTs were failed at the most conservative cutoff). An EI-5 ≤ 1 is classified as an overall *Pass*, in that it contains at most one failure at the most liberal cutoff (Erdodi, 2019).

The interpretation of EI-5 values 2 and 3 is more problematic, for they can indicate either a single failure at a more conservative cutoff, or multiple failures at the most liberal cutoff. Regardless of the specific configuration, this range of performance fails to deliver sufficiently strong evidence to render the entire response set *invalid*. At the same time, it signals subthreshold evidence of non-credible responding (Abeare, Messa, Whitfield, et al., 2018a; Erdodi, Hurtubise, et al., 2018c; Erdodi, Seke, Shahein, et al., 2017c; Erdodi et al., 2018f; Proto et al., 2014).

Therefore, this range of performance on the EI-5 is labeled *Borderline* and excluded from analyses requiring a dichotomous (Pass/Fail) criterion variable (Erdodi, 2019). However, an EI-5 ≥ 4 indicates either at least two failures at the more conservative cutoffs, at least four failures at the most liberal cutoff, or some combination of both. As such, they meet commonly used standards for classifying the entire neurocognitive profile *invalid* (Boone, 2013; Davis & Millis, 2014; Larrabee, 2014; Odland, Lammy, Martin, Grote, & Mittenberg, 2015). The majority of the sample produced a score in the passing range on both versions of the EI-5, with a roughly 20% BR$_{Fail}$ (Table 2).

Given that the EI-model is a relatively new approach to multivariate performance validity assessment, it was validated against the WMT, a well-stablished free-standing PVT (Green, Iverson, & Allen, 1999; Iverson, Green, & Gervais, 1999; Tan, Slick, Strauss, & Hultsch, 2002), in order to demonstrate its diagnostic utility within the present sample. The EI-5$_{AWM}$ produced a significantly higher (.91) overall classification accuracy than the EI-5$_{PSP}$ (.79), driven by superior sensitivity (.74 vs. .47). However, both versions of the EI-5 were

Table 1 Individual components of the EI-5s and base rates of failure at given cutoffs ($n = 100$)

	EI-5 Values					EI-5 Values			
EI-5$_{PSP}$	**0**	**1**	**2**	**3**	**EI-5$_{AWM}$**	**0**	**1**	**2**	**3**
Components	Pass	Fail	FAIL	**FAIL**	Components	Pass	Fail	FAIL	**FAIL**
CD$_{WAIS-IV}$	>5	5	4	≤3	FCR$_{CVLT-II}$	16	15	14	≤13
Base Rate	*75.0*	*13.0*	*6.0*	*6.0*	*Base Rate*	*75.0*	*7.0*	*8.0*	*10.0*
GPB$_{DH}$	>29	24-29	21-23	≤20	LM$_{WMS-IV}$	>20	19-20	17-18	≤16
Base Rate	*80.0*	*10.0%*	*4.%*	*6.0%*	*Base Rate*	*81.0*	*10.0*	*4.0*	*5.0*
OMI$_{CPT-II}$	≤70	71-149	150-200	>200	LNS$_{WAIS-IV}$	>7	7	6	≤5
Base Rate	*75.0*	*16.0*	*5.0*	*5.0*	*Base Rate*	*75.0*	*13.0*	*6.0*	*6.0*
SS$_{WAIS-IV}$	>6	5-6	3-4	≤2	RDS	>7	7	6	≤5
Base Rate	*71.0*	*17.0*	*7.0*	*5.0*	*Base Rate*	*76.0*	*16.0*	*4.0*	*5.0*
TMT-A	>37	26-37	22-25	≤21	VPA$_{WMS-IV}$	>34	30-34	28-29	≤27
Base Rate	*72.0*	*15.0*	*7.0*	*5.0*	*Base Rate*	*76.0*	*14.0*	*4.0*	*6.0*

Base rates are italicized for easier visual separation. Shading reflects the level of confidence in correctly classifying ascore in a given range as invalid: a darker shade represents increased likelihood of non-credible responding. *EI-5$_{PSP}$* Erdodi Index Five – Processing speed; *EI-5$_{AWM}$* Erdodi Index Five – Attention, working memory and auditory verbal memory; *CD$_{WAIS-IV}$* Coding subtest of the Wechsler Adult Intelligence Scale – Fourth Edition (age-corrected scaled score; Erdodi, Abeare et al., 2017a, 2018a; Etherton, Bianchini, Heinly, & Greve, 2006; Inman & Berry, 2002; Kim et al., 2010; Trueblood, 1994); *GPB$_{DH}$* Grooved Pegboard Test demographically adjusted T-score (Erdodi, Kirsch, et al., 2018c; Erdodi, Seke, Shahein, et al., 2017c); *OMI$_{CPT-II}$* omission errors on the Conners' Continuous Performance Test – Second Edition (T-scores; Erdodi et al., 2014b; Erdodi, Pelletier, & Roth, 2018e; Lange et al., 2013; Ord et al., 2010); *SS$_{WAIS-IV}$* Symbol Search subtest of the Wechsler Adult Intelligence Scale – Fourth Edition (age-corrected scaled score; Erdodi, Abeare, Lichtenstein et al., 2017; Etherton et al., 2006); *TMT-A* Trail Making Test – Part A (T-score; Ashendorf et al., 2017; Erdodi & Lichtenstein, 2019); *FCR$_{CVLT-II}$* The Forced Choice Recognition trial of the California Verbal Learning Test – Second Edition (raw score; Erdodi, Abeare, Medoff, et al., 2017b; Persinger et al., 2018; Schwartz et al., 2016); *LM$_{WMS-IV}$* WMS-IV Logical Memory Recognition (raw score; Bortnik et al., 2010; Pearson, 2009); *LNS$_{WAIS-IV}$* Letter-Number Sequencing subtest of the Wechsler Adult Intelligence Scale – Fourth Edition (age-corrected scaled score; Shura, Miskey, Rowland, Yoash-Gatz, & Denning, 2016); *RDS* reliable digit span (Babikian, Boone, Lu, & Arnold, 2006; Greiffenstein, Baker, & Gola, 1994; Pearson, 2009; Reese, Suhr, & Riddle, 2012); *VPA$_{WMS-IV}$* WMS-IV Verbal Paired Associates Recognition (raw score; Pearson, 2009); *VR$_{WMS-IV}$* WMS-IV Visual Reproduction Recognition (raw score; Pearson, 2009)

highly specific (.94) to psychometrically defined non-credible responding (Table 3).

Procedure

Patients were assessed in two half-day (4-hour) appointments in an outpatient setting. Psychometric testing was performed by Master's level psychologists who received specialized training in test administration and scoring. The clinical interview and the interpretation of neuropsychological profiles were performed by a staff clinical neuropsychologist (fifth author), who was also responsible for generating an integrative summary report. By design, this is a retrospective archival study. Clinical data were fully and irreversibly de-identified prior to being used for research purposes. The project was approved by the institutional board overseeing compliance with research ethics.

Data Analysis

Descriptive statistics (mean, standard deviation, range, BR$_{Fail}$) were reported when relevant. Overall classification accuracy

(AUC) and the corresponding 95% confidence intervals were computed using SPSS version 23.0. An AUC in the .70–.79 range is classified as *acceptable*; a value in the .80–.89 range is classified as *excellent*, whereas an AUC ≥ .90 is considered *outstanding* (Hosmer & Lemeshow, 2000). Sensitivity and specificity were computed using standard formulas (Grimes & Schulz, 2005). In the context of PVTs, the minimum acceptable threshold for specificity is .84 (Larrabee, 2003), although values ≥ .90 are desirable and are becoming the emerging norm (Boone, 2013; Donders & Strong, 2011). The statistical significance of risk ratios (RR) was determined using the chi-square test of independence.

Results

Dominant Hand T-Score Cutoffs

Dominant hand FTT T-scores produced significant AUCs (.70–.81) against all three criterion PVTs (Table 4). A cutoff of $T \leq 29$ was highly specific (.97–1.00) to psychometrically

Table 2 Frequency, cumulative frequency, and classification range for the first ten levels of the EI-5s

EI-5	$EI\text{-}5_{PSP}$		$EI\text{-}5_{AWM}$		Classification	
	f	Cumulative %	f	Cumulative %	By row	Overall
0	47	47.0	44	44.0	PASS	PASS
1	19	66.0	14	58.0	Pass	
2	9	75.0	9	67.0	Borderline	
3	4	79.0	10	77.0	Borderline	
4	3	82.0	8	85.0	Fail	FAIL
5	4	86.0	3	88.0	Fail	
6	3	89.0	4	92.0	FAIL	
7	3	92.0	2	94.0	FAIL	
8	0	92.0	1	95.0	**FAIL**	
9	1	93.0	1	96.0	**FAIL**	

Shading reflects the level of confidence in correctly classifying a score in a given range as invalid: a darker shade represents increased likelihood of non-credible responding. $EI\text{-}5_{PSP}$ Erdodi Index Five – Processing speed; $EI\text{-}5_{AWM}$ Erdodi Index Five – Attention, working memory and auditory verbal memory

defined non-credible responding, but had variable sensitivity (.23–.45). Increasing the cutoff to ≤ 33 resulted in a predictable trade-off between improved sensitivity (.36–.55) and diminished, but still high specificity (.90–.98). Making the cutoff even more liberal (≤ 35) reached the point of diminished return: a loss in specificity (.87–.94) without any corresponding gain in sensitivity.

Non-dominant Hand T-Score Cutoffs

Non-dominant hand FTT T-scores produced significant AUCs (.67–.74) against all three criterion PVTs. A cutoff of $T \leq 29$ had uniformly good specificity (.92–.93), but small and variable sensitivity (.23–.44). Increasing the cutoff to ≤ 33 was largely inconsequential to both sensitivity (.28–.44) and specificity (.91–.92). No patient scored $T = 35$. Making the cutoff even more liberal (≤ 37) resulted in a meaningful increase in sensitivity (.41–.50). However, it also caused specificity to drop below the minimum threshold (.83) against the $EI\text{-}5_{PSP}$, while maintaining adequate specificity against the other two criterion PVTs (.88–.90).

Combined T-Score Cutoffs

Consistent with the practice established in previous research on measures of manual dexterity as a PVT (Arnold et al., 2005; Axelrod et al., 2014; Erdodi, Seke, Shahein, et al., 2017c; Larrabee, 2003), a third variable combining both trials (COM) was created. *Fail* on COM was defined as failing a given cutoff with *both* hands; *Pass* was defined as passing a given cutoff with *either* hand. As such, failing the COM provides a more conservative index of non-credible responding, because it allows a single failure to be counted as an overall *Pass*.

A COM cutoff of $T \leq 33$ had very high specificity (.97–.98), but low sensitivity (.18–.33). Increasing the cutoff to ≤ 37 preserved the specificity (.97–.98) against the WMT and $EI\text{-}5_{AWM}$, while improving sensitivity (.36–.43). However, this cutoff disproportionately traded specificity (.89) for sensitivity (.43) against the $EI\text{-}5_{PSP}$. Making the cutoff even more liberal (≤ 39) resulted in negligible changes in classification accuracy: It further deflated specificity (.86–.95) and produced comparable gains in sensitivity (.38–.50).

Table 3 Classification accuracy of the EI-5s against the WMT as criterion PVT

| EI Version | n | Cutoff | BR_{Fail} | Green's Word Memory Test | | | |
| | | | | Standard *Pass/Fail* | | | |
				AUC	95% CI	39.0% Sensitivity	Specificity
$EI\text{-}5_{PSP}$	87	≥ 4	21.0	.79	.71–.88	.47	.94
$EI\text{-}5_{AWM}$	81	≥ 4	23.0	.91	.85–.97	.74	.94

$EI\text{-}5_{PSP}$ Erdodi Index Five – Processing speed; $EI\text{-}5_{AWM}$ Erdodi Index Five – Attention, working memory and auditory verbal memory; *PVT* performance validity test; *AUC* area under the curve; BR_{Fail} base rate of failure

Table 4 Base rates of failure and classification accuracy of FTT validity cutoffs against the WMT and the EI-5s as criterion PVTs

Hand	Scale	BR$_{Fail}$		WMT n = 100 SENS 39.0%	SPEC	EI-5$_{PSP}$ n = 87 SENS 21.4%	SPEC	EI-5$_{AWM}$ n = 81 SENS 28.4%	SPEC
			AUC	.70		.77		.81	
	T		95% CI	*.59–.81*		*.64–.91*		*.57–.84*	
DOM	≤ 29	10.0%		.26	.97	.23	1.00	.45	1.00
	≤ 31	–		–	–	–	–	–	–
	≤ 33	15.0%		.36	.90	.40	.98	.55	.90
	≤ 35	20.0%		.36	.87	.40	.94	.55	.89
	≤ 37	24.0%		.41	.83	.47	.91	.55	.83
			AUC	.67		.74		.66	
	T		95% CI	*.56–.78*		*.60–.88*		*.52–.81*	
ND	≤ 29	14.0%		.23	.92	.44	.92	.30	.93
	≤ 31	–		–	–	–	–	–	–
	≤ 33	16.0%		.28	.92	.44	.91	.35	.91
	≤ 35	–		–	–	–	–	–	–
	≤ 37	22.0%		.41	.90	.50	.83	.48	.88
	T								
COM	≤ 33	9.0%		.21	.98	.39	.97	.30	.98
	≤ 35	–		–	–	–	–	–	–
	≤ 37	15.0%		.36	.98	.44	.89	.43	.97
	≤ 39	18.0%		.38	.95	.50	.86	.48	.93
	≤ 41	25.0%		.44	.87	.67	.80	.48	.84
			AUC	.70		.75		.74	
	Raw		95% CI	*.59–.81*		*.62–.89*		*.61–.87*	
DOM	≤ 28/35	6.0%		.13	.98	.33	1.00	.22	.98
			AUC	.67		.70		.71	
	Raw		95% CI	*.55–.78*		*.54–.86*		*.58–.85*	
COM	≤ 58/66	5.0%		.13	1.00	.22	.98	.17	1.00

Raw score cutoffs are gender specific (women/men)

FTT Finger Tapping Test demographically adjusted T-score; *WMT* Green's Word Memory Test (standard cutoffs); *PVT* performance validity test; *EI-5$_{FCR}$* Erdodi Index Five – Forced choice recognition (*Fail* ≥ 4); *EI-5$_{PSP}$* Erdodi Index Five – Processing speed; *EI-5$_{AWM}$* Erdodi Index Five – Attention, working memory and auditory verbal memory; *AUC* area under the curve; *DH* dominant hand; *ND* non-dominant hand; *COM* combined FTT scores (*Fail* defined as scoring at or below the cutoff on *both* hands; *Pass* defined as a score above the cutoff with *either* hand); *BR$_{Fail}$* base rate of failure; *SENS* sensitivity; *SPEC* specificity

Raw Score Cutoffs

To provide a head-to-head comparison with T-score-based cutoffs, the classification accuracy of raw score cutoffs has also been computed. Dominant hand raw score cutoffs produced significant AUCs (.70–.75) against all three criterion PVTs. Gender-corrected cutoffs (≤ 28 for women and ≤ 35 for men; Arnold et al., 2005) had low BR$_{Fail}$ (6.0%). Consequently, they resulted in high specificity (.98–1.00) and low sensitivity (.13–.33). Combined hand raw score cutoffs produced similar signal detection profile: low BR$_{Fail}$ (5.0%), high specificity (.98–1.00), and low sensitivity (.13–.22).

Relationship Between Injury Severity and Failing FTT Validity Cutoffs

Since severe TBI could result in decreased motor speed (Curtis, Greve, & Bianchini, 2009; Donders & Strong, 2015; Erdodi, Abeare, et al., 2017a, 2018a; Haaland, Temkin, Randahl, & Dikmen, 1994) or difficulty inhibiting the

Table 5 Base rate of failure (%) as a function of injury severity across various PVTs and cutoffs

		Injury severity				
		$n = 74$	$n = 26$			
PVT	Cutoff	Mild	M/S	RR	χ^2	p
WMT	Standard	47.3	15.4	3.07	8.24	.004
EI-5$_{PSP}$	≥ 4	25.8	9.1	2.84	2.69	.101
EI-5$_{AWM}$	≥ 4	36.7	4.8	7.65	7.79	.005
FTT$_{DH}$	≤ 29	10.8	7.7	1.40	0.21	.648
	≤ 33	16.2	11.5	1.41	0.33	.566
FTT$_{ND}$	≤ 29	16.2	7.7	2.10	1.16	.281
	≤ 33	18.9	7.7	2.45	1.80	.179
FTT$_{COM}$	≤ 33	10.8	3.8	2.84	1.14	.286
	≤ 37	18.9	3.8	4.97	3.43	.064

FTT Finger Tapping Test demographically adjusted T-score; *WMT* Green's Word Memory Test (standard cutoffs); *PVT* performance validity test; *EI-5$_{PSP}$* Erdodi Index Five – Processing speed; *EI-5$_{AWM}$* Erdodi Index Five – Attention, working memory and auditory verbal memory; *DH* dominant hand; *ND* non-dominant hand; *COM* combined FTT scores (*Fail* defined as scoring at or below the cutoff on *both* hands; *Pass* defined as a score above the cutoff with *either* hand); *M/S* moderate-to-severe; *RR* risk ratio

concurrent movement of other fingers during tapping (Prigatano & Borgaro, 2003), it is important to ensure that genuine impairment is not misclassified as non-credible responding. Therefore, the BR$_{Fail}$ was compared as a function

of injury severity. Mild TBI was invariably associated with higher BR$_{Fail}$ on all PVTs examined (RR 1.40–7.65) compared to moderate-to-severe TBI (Table 5). The contrasts involving the WMT and the EI-5$_{AWM}$ were statistically significant ($p < .01$).

Old and New: A Head-to-Head Comparison of Raw and T-Score Cutoffs

Finally, FTT validity cutoffs based on raw scores were compared to demographically adjusted ones in order to empirically evaluate their relationship with age, level of education, and sex. No significant age effects emerged as a function of passing or failing either type of cutoff (Table 6). However, a large effect ($d = 1.00$) was observed among the subset of patients who failed the COM raw score cutoff. BR$_{Fail}$ was evenly distributed between males and females using T-score-based cutoffs (RR 0.85–1.27; p values .598–.821).

However, women failed raw score cutoffs at higher rates than men (RR 1.89–2.53), although the difference in failure rate was not statistically significant (p values .249–1.89). It must be pointed out that the sample size for the *Fail* group was notably smaller than that of the *Pass* group for both the T-score (n: 15–16 vs. 84–85) and raw score-based (n: 5–6 vs. 95–94) cutoffs. Therefore, statistical tests were likely underpowered by unequal group and small overall sample size (Table 7).

Table 6 Age and level of education as a function of passing or failing various FTT validity cutoffs

FTT variable					Demographic variables									
					Age (years)					Education (years)				
Hand	Score	Cutoff		n	M	SD	t	p	d	M	SD	t	p	d
DOM	T	≤ 33	Pass	85	38.7	15.4	0.14	.886	–	13.7	2.7	0.47	.643	–
			Fail	15	39.3	12.2				13.4	2.4			
ND	T	≤ 33	Pass	84	38.8	15.2	0.02	.984	–	13.9	2.7	1.48	.143	–
			Fail	16	38.8	13.9				12.8	1.5			
COM	T	≤ 37	Pass	85	38.4	15.5	0.69	.493	–	13.9	2.7	1.44	.153	–
			Fail	15	41.3	10.5				12.8	1.6			
DOM	Raw	≤ 28[a]	Pass	94	38.7	15.0	0.34	.734	–	13.7	2.6	0.51	.615	–
		≤ 35[b]	Fail	6	40.8	14.0				13.2	3.5			
COM	Raw	≤ 58[a]	Pass	95	38.5	14.9	0.86	.393	–	13.8	2.6	1.68	.097	1.00
		≤ 66[b]	Fail	5	44.4	15.7				11.8	1.1			

FTT Finger Tapping Test; *DOM* dominant; *ND* non-dominant; *COM* combined (*Fail* defined as scoring at or below the cutoff on *both* hands; *Pass* defined as a score above the cutoff with *either* hand)

[a] Women

[b] Men

Italics improve the visual discrimination of important details

Table 7 Base rate of failure (%) in males and females across various FTT validity cutoffs

			Sex				
FTT variable			Male	Female			
Hand	Score	Cutoff	$n = 56$	$n = 44$	RR	χ^2	p
DOM	T	≤ 33	14.3	15.9	1.11	0.05	.821
ND	T	≤ 33	14.3	18.2	1.27	0.28	.598
COM	T	≤ 37	16.1	13.6	0.85	0.12	.735
DOM	Raw	≤ 28[a]	3.6	9.1	2.53	1.33	.249
		≤ 35[b]					
COM	Raw	≤ 58[a]	3.6	6.8	1.89	0.55	.460
		≤ 66[b]					

FTT Finger Tapping Test; *DOM* dominant; *ND* non-dominant; *COM* combined (*Fail* defined as scoring at or below the cutoff on *both* hands; *Pass* defined as a score above the cutoff with *either* hand); *RR* risk ratio

[a] Women

[b] Men

Discussion

This study was designed to compare the classification accuracy of raw score-based and demographically adjusted validity cutoffs embedded within the FTT. We predicted that the two types of cutoffs (raw vs. T-scores) would have comparable signal detection performance. While this was the case in terms of AUC values, demographically adjusted validity cutoffs had notably higher (i.e., roughly double) sensitivity than raw score-based cutoffs at comparable levels of specificity. In other words, raw score-based cutoffs disproportionally sacrificed their ability to detect non-credible performance in order to ensure a low false positive rate.

Our second hypothesis had mixed support. We predicted that T-score-based FTT validity cutoffs would better protect demographically disadvantaged (i.e., female, older, less educated) examinees against the threat of false positive errors. Surprisingly, both sets of cutoffs were equally robust to age effects. Failing combined hand raw score cutoffs was indeed associated with lower levels of education, whereas no difference was observed across T-score cutoffs. Paradoxically, women failed the gender-corrected raw score cutoffs introduced by Arnold et al. (2005) roughly twice as often as men. Again, no sex difference was observed across T-score cutoffs.

Taken together, these findings suggest that demographically adjusted cutoffs not only have superior classification accuracy but also provide a better balance between sensitivity and specificity while neutralizing the potential confounding effects of age, sex, and level of education. By pre-emptively accounting for these demographic variables, assessors improve their diagnostic accuracy in clinical

settings and protect themselves against challenges in medico-legal settings (i.e., avoid the perception of unchecked biases). Logistically, it is also easier to apply uniform cutoffs rather than unique ones for different demographic categories and dominant vs. non-dominant hand on the FTT. These advantages may explain the growing trend towards demographically adjusted cutoffs (Ashendorf, Clark, & Sugarman, 2017; Erdodi, Kirsch, et al., 2018d; Erdodi & Lichtenstein, 2019).

The modality specificity effect (Erdodi, 2019; Rai & Erdodi, 2019) was observed during both the cross-validation of the EI-5s and the FTT cutoffs. The EI-5$_{AWM}$, the modality congruent predictor variable, produced significantly higher AUC value against the WMT, the criterion PVT. Although its specificity was identical to that of the modality incongruent counterpart (EI-5$_{PSP}$), sensitivity was notably higher (.47 vs. .74). Similarly, FTT cutoffs tended to produce the best classification accuracy against the EI-5$_{PSP}$, the modality congruent criterion. This engineered method variance was introduced to evaluate the signal detection profile of FTT validity cutoffs across changing definitions of invalid performance, consistent with the methodological triangulation proposed by Campbell and Fiske (1959). The fact that the optimal cutoffs (dominant and non-dominant hand $T \leq 33$ as well as combined ≤ 37) maintained high levels of specificity regardless of the criterion PVT further consolidates the evidence base supporting their clinical utility .

Equally important, the newly introduced FTT validity cutoffs were insensitive to injury severity—an important feature of a good PVT (Donders, 2005). In fact, patients with mild TBI produced consistently higher BR$_{Fail}$ on the FTT than those with moderate or severe TBI. This paradoxical finding was even more pronounced on the WMT and the EI-5$_{AWM}$, and is well-replicated in the research literature (Carone, 2008; Erdodi & Rai, 2017; Green, Flaro, & Courtney, 2009; Green et al., 1999; Ord, Boettcher, Greve, & Bianchini, 2010; Sweet, Goldman, & Guidotti Breting, 2013). This pattern of reverse dose-response relationship (Hill, 1965) provides additional evidence that PVT failures in mild TBI are unlikely to be attributable to genuine neurological impairment (Critchfield et al., 2019).

Consistent with previous reports (Arnold et al., 2005; Axelrod et al., 2014; Rapport et al., 1998), dominant hand FTT validity cutoffs provided a better overall measure of the credibility of a given response set (i.e., higher classification accuracy, more stable parameter estimates) compared to non-dominant hand FTT validity cutoffs. Therefore, if one must select one validity cutoff, our results suggest that a dominant hand T-score ≤ 33 is the single best demarcation line between credible and non-credible responding on the FTT.

Inevitably, the study also has a number of limitations, the most obvious of which is the relatively small sample size collected from a single geographic region. Consequently, the distribution of FTT scores was truncated, and certain T-scores were not observed at all (Table 4). Therefore, the results may be vulnerable to sample-specific findings and should be replicated before applying them broadly to clinical and forensic practice (Lichtenstein et al., 2019). Also, since all patients in this study were evaluated for residual deficits subsequent to TBI, it is unclear whether the present findings would generalize to different clinical populations. Future studies on larger samples with more diverse etiologies are needed to establish confidence in the proposed FTT validity cutoffs.

However, the study also has a number of strengths. It expanded the existing knowledge base on raw score-based FTT cutoffs (Arnold et al., 2005; Larrabee, 2003; Axelrod et al., 2014) by introducing demographically adjusted validity cutoffs, following a growing trend in EVI development (Ashendorf et al., 2017; Brooks & Ploetz, 2015; Erdodi & Lichtenstein, 2019; Erdodi, Lichtenstein, Rai, & Flaro, 2017b; Erdodi, Pelletier, & Roth, 2018e; Erdodi et al., 2016; Kirkwood, Hargrave, & Kirk, 2011; Lichtenstein, Flaro, Baldwin, Rai, & Erdodi, 2019; Sussman, Peterson, Connery, Baker, & Kirkwood, 2019). In addition, it also provided further validation to the EI-5 as a multivariate model of performance validity assessment.

Compliance with Ethical Standards

All data collection, storage, and processing were done in compliance with the Helsinki Declaration.

Conflict of Interest Drs. Erdodi and Abeare are employed by the University of Windsor. In addition to that, they provide forensic consultation and medicolegal assessments, for which they receive financial compensation.

Human and Animal Rights and Informed Consent Relevant ethical guidelines regulating research involving human participants were followed throughout the project.

References

Abeare, C., Messa, I., Whitfield, C., Zuccato, B., Casey, J., Rykulski, N., & Erdodi, L. (2018a). Performance validity in collegiate football athletes at baseline neurocognitive testing. *The Journal of head trauma rehabilitation*, https://doi.org/10.1097/HTR.000000000000451.

Abeare, C. A., Messa, I., Zuccato, B. G., Merker, B., & Erdodi, L. A. (2018b). 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.

Allen, L., & Green, P. (1999). CARB and WMT effort test scores in 57 patients with severe traumatic brain injury. *Archives of Clinical Neuropsychology, 14*(8), 789.

American Congress on Rehabilitation Medicine. (1993). Definition of mild traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 8*(3), 86–87.

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.

An, K. Y., Charles, J., Ali, S., Enache, A., Dhuga, J., & Erdodi, L. A. (2019). Reexamining performance validity cutoffs within the complex ideational material and the Boston naming test–short form using an experimental malingering paradigm. *Journal of Clinical and Experimental Neuropsychology, 41*(1), 15–25.

Armistead-Jehle, P., & Denney, R. L. (2015). The detection of feigned impairment using the WMT, MSVt, and NV-MSVT. *Applied Neuropsychology. Adult, 22*(2), 147–155. https://doi.org/10.1080/23279095.2014.880842.

Arnold, G., Boone, K. B., Lu, P., Dean, A., Wen, J., Nitch, S., & McPhearson, S. (2005). Sensitivity and specificity of finger tapping test scores for the detection of suspect effort. *The Clinical Neuropsychologist, 19*(1), 105–120. https://doi.org/10.1080/13854040490888567.

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.

Axelrod, B. N., Meyers, J. E., & Davis, J. J. (2014). Finger tapping test performance as a measure of performance validity. *The Clinical Neuropsychologist, 28*(5), 876–888. https://doi.org/10.1080/13854046.2014.907583.

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*(1), 145–159.

Baker, D. A., Connery, A. K., Kirk, J. W., & Kirkwood, M. W. (2014). Embedded performance validity indicators within the California verbal learning test, Children's version. *The Clinical Neuropsychologist, 28*(1), 116–127.

Bauer, L., Yantz, C. L., Ryan, L. M., Warden, D. L., & McCaffrey, R. J. (2005). An examination of the California verbal learning test II to detect incomplete effort in a traumatic brain injury sample. *Applied Neuropsychology, 12*(4), 202–207. https://doi.org/10.1207/s15324826an1204_3.

Bigler, E. D. (2012). Symptom validity testing, effort and neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 632–642. https://doi.org/10.1017/S1355617712000252.

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

Boone, K. B. (2013). *Clinical practice of forensic neuropsychology*. New York: Guilford.

Bortnik, K. E., Boone, K. B., Marion, S. D., Amano, S., Ziegler, E., Victor, T. L., & Zeller, M. A. (2010). Examination of various WMS-III logical memory scores in the assessment of response bias.

The Clinical Neuropsychologist, 24(2), 344–357. https://doi.org/10.1080/13854040903307268.

Brooks, B. L., & Ploetz, D. M. (2015). Embedded performance validity on the CVLT-C for youth with neurological disorders. *Archives of Clinical Neuropsychology, 30*, 200–206.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity (NAN policy and planning committees). *Archives of Clinical Neuropsychology, 20*, 419–426.

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.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105.

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*, 960–971.

Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., Larrabee, G. J., & Ord, J. S. (2015). Official position of the American Academy of clinical neuropsychology Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist, 29*(6), 723–740.

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.

Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations. (2015). Cognitive Tests and Performance Validity Tests. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK305230/

Conners, K. C. (2004). *Conner's Continuous Performance Test (CPT II). Version 5 for Windows. Technical Guide and Software Manual*. North Tonawada: Multi-Health Systems.

Costabile, T., Bilo, L., DeRosa, A., Pane, C., & Sacca, F. (2018). Dissociative identity disorder: Restoration of executive functions after switch from alter to host personality. *Psychiatry and Clinical Neurosciences, 72*, 189–190.

Cottingham, M. E., Victor, T. L., Boone, K. B., Ziegler, E. A., & Zeller, M. (2014). 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.

Critchfield, E. A., Soble, J. R., Marceaux, J. C., Bain, K. M., Bailey, K. C., et al. (2018). Cognitive impairment does not cause invalid performance: Analyzing performance patterns among cognitively unimpaired, impaired, and noncredible participants across six performance validity tests. *The Clinical Neuropsychologist*, https://doi.org/10.1080/13854046.2018.1508615.

Curtis, K. L., Greve, K. W., & Bianchini, K. J. (2009). The Wechsler adult intelligence scale-III and malingering in traumatic brain injury. *Assessment, 16*(4), 401–414.

Davis, J. J., & Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist, 28*(2), 199–214. https://doi.org/10.1080/13854046.2014.884633.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. (2000). *The California Verbal Learning Test* (Second ed.). San Antonio: The Psychological Corporation.

Donders, J. (2005). Performance on the test of memory malingering in a mixed pediatric sample. *Child Neuropsychology, 11*(2), 221–227.

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.

Donders, J., & Strong, C. H. (2013). Does Green's word memory test really measure memory? *Journal of Clinical and Experimental Neuropsychology, 35*(8), 827–834.

Donders, J., & Strong, C. H. (2015). Clinical utility of the Wechsler adult intelligence scale – Fourth edition after traumatic brain injury. *Assessment, 22*(1), 17–22.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (Fourth ed.). San Antonio: Pearson.

Erdodi, L. A. (2019). Aggregating validity indicators: The salience of domain specificity and the indeterminate range in multivariate models of performance validity assessment. *Applied Neuropsychology. Adult, 26*(2), 155–172. https://doi.org/10.1080/23279095.2017.1384925.

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., & Rai, J. K. (2017). A single error is one too many: Examining alternative cutoffs on trial 2 of 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., Kirsch, N. L., Lajiness-O'Neill, R., Vingilis, E., & Medoff, B. (2014a). 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.

Erdodi, L. A., Roth, R. M., Kirsch, N. L., Lajiness-O'Neill, R., & Medoff, B. (2014b). Aggregating validity indicators embedded in Conners' CPT-II outperforms individual cutoffs at separating valid from invalid performance in adults with traumatic brain injury. *Archives of Clinical Neuropsychology, 29*(5), 456–466. https://doi.org/10.1093/arclin/acu026.

Erdodi, L. A., Tyson, B. T., Abeare, C. A., Lichtenstein, J. D., Pelletier, C. L., Rai, J. K., & Roth, R. M. (2016). The BDAE complex ideational material – A measure of receptive language or performance validity? *Psychological Injury and Law, 9*, 112–120. https://doi.org/10.1007/s12207-016-9254-6.

Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B. T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017a). 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., Lichtenstein, J. D., Rai, J. K., & Flaro, L. (2017b). 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., Seke, K. R., Shahein, A., Tyson, B. T., Sagar, S., & Roth, R. M. (2017c). Low scores on the grooved pegboard test are associated with invalid responding and psychiatric symptoms. *Psychology & Neuroscience, 10*(3), 325–344. https://doi.org/10.1037/pne0000103.

Erdodi, L. A., Abeare, C. A., Medoff, B., Seke, K. R., Sagar, S., & Kirsch, N. L.(2018a). 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.

124

Psychol. Inj. and Law (2019) 12:113–126

Erdodi, L. A., Dunn, A. G., Seke, K. R., Charron, C., McDermott, A., Enache, A., Maytham, C., & Hurtubise, J. L. (2018b). The Boston naming test as a measure of performance validity. *Psychological Injury and Law, 11*(1), 1–8.

Erdodi, L. A., Hurtubise, J. L., Charron, C., Dunn, A., Enache, A., McDermott, A., & Hirst, R. (2018c). The D-KEFS trails as performance validity tests. *Psychological Assessment, 30*(8), 1081–1095.

Erdodi, L. A., Kirsch, N. L., Sabelli, A. G., & Abeare, C. A. (2018d). 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., Pelletier, C. L., & Roth, R. M. (2018e). 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., Tyson, B. T., Abeare, C. A., Zuccato, B. G., Rai, J. K., Seke, K. R., Sagar, S., & Roth, R. M. (2018f). 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.

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. https://doi.org/10.1080/13803390500346595.

Gladsjo, J. A., Schuman, C. C., Evans, J. D., Peavy, G. M., Miller, S. W., & Heaton, R. K. (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment, 6*(2), 147–178.

Green, P. (2003). *Green's word memory test*. Edmonton: Green's Publishing.

Green, P., & Flaro, L. (2003). Word memory test performance in children. *Child Neuropsychology, 9*(3), 189–207.

Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malingering in head injury litigation with the word memory test. *Brain Injury, 13*(10), 813–819. https://doi.org/10.1080/026990599121205.

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*, 741–750.

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.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and post-concussion syndrome. *Journal of the International Neuropsychological Society, 2*(6), 477–485.

Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the test of memory malingering in traumatic brain injury: Results of a known-group analysis. *Journal of Clinical and Experimental Neuropsychology, 28*(7), 1176–1190.

Greve, K. W., Etherton, J. L., Ord, J., Bianchini, K. J., & Curtis, K. L. (2009). Detecting malingered pain-related disability: Classification accuracy of the test of memory malingering. *The Clinical Neuropsychologist, 16*(2), 179–191.

Grimes, D. A., & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *The Lancet, 365*(9469), 1500–1505.

Guilmette, T.J. (2013). The role of clinical judgment in symptom validity testing. In D. A. Carone S. S. Bush (Eds.), *Mild traumatic injury: Symptom validity assessment and malingering* (pp. 31–43). New York, NY: Springer.

Haaland, K. Y., Temkin, N., Randahl, G., & Dikmen, S. (1994). Recovery of simple motor skills after head injury. *Journal of Clinical and Experimental Neuropsychology, 16*(3), 448–456.

Halstead, W. (1947). *Brain and intelligence. A quantitative study of the frontal lobes*. Chicago: University of Chicago Press.

Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*(5), 892–900.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan Battery*. Odessa: Psychological Assessment Resources.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). *Wisconsin Card Sorting Test (WCST) manual revised and expanded*. Odessa: Psychological Assessment Resources.

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.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants 1. (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.

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.

Homann, C. N., Quehenberger, F., Petrovic, K., Hartung, H. P., Ruzicka, E., Homann, B., et al. (2003). Influence of age, gender, education and dexterity on upper limb motor performance in parkinsonian patients and healthy controls. *Journal of Neural Transmission, 110*, 885–897. https://doi.org/10.1007/s00702-003-00097.

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

Inman, T. H., & Berry, D. T. R. (2002). Cross-validation of indicators of malingering: A comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology, 17*, 1–23.

Iverson, G., Green, P., & Gervais, R. (1999). Using the word memory test to detect biased responding in head injury litigation. *Journal of Cognitive Rehabilitation, 17*(2), 4–8.

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.

Kalfon, T. B., Gal, G., Shorer, R., & Ablin, J. N. (2016). Cognitive functioning in fibromyalgia: The central role of effort. *Journal of Psychosomatic Research, 87*, 30–36.

Kim, N., Boone, K. B., Victor, T., Lu, P., Keatinge, C., & Mitchell, C. (2010). Sensitivity and specificity of a digit symbol recognition trial in the identification of response bias. *Archives of Clinical Neuropsychology, 25*(5), 420–428. https://doi.org/10.1903/arclin/acq040.

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV digit span subtest in detecting noncredible effort performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26*, 377–384.

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.

Lange, R. T., Iverson, G. L., Brickell, T. A., Staver, T., Pancholi, S., Bhagwat, A., & French, L. M. (2013). Clinical utility of the Conners' continuous performance test-II to detect poor effort in U.S. military personnel following traumatic brain injury. *Psychological Assessment, 25*(2), 339–352.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*, 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*, 410–425. https://doi.org/10.1080/13854040701494987.

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). Minimizing false positive errors with multiple performance validity tests: Response to Bilder, sugar, and Hellemann. *The Clinical Neuropsychologist, 28*(8), 1230–1242.

Lichtenstein, J. D., Erdodi, L. A., & Linnea, K. S. (2017). Introducing a forced-choice recognition task to the California verbal learning test – Children's version. *Child Neuropsychology, 23*(3), 284–299. https://doi.org/10.1080/09297049.2015.1135422.

Lichtenstein, J. D., Erdodi, L. A., Rai, J. K., Mazur-Mosiewicz, A., & Flaro, L. (2018a). Wisconsin card sorting test embedded validity indicators developed for adults can be extended to children. *Child Neuropsychology, 24*(2), 247–260. https://doi.org/10.1080/09297049.2016.1259402.

Lichtenstein, J. D., Holcomb, M., & Erdodi, L. A. (2018b). 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*(3–4), 94–104.

Lichtenstein, J. D., Flaro, L., Baldwin, F., Rai, J. K., & Erdodi, L. A. (2019). Further evidence for embedded validity tests in children within the Conners' continuous performance test – Second edition. *Developmental Neuropsychology, 44*(2), 159–171. https://doi.org/10.1080/87565641.2019.1565536.

Lichtenstein, J. D., Greenacre, M. K., Cutler, L., Abeare, K., Baker, S. D., Kent, K., J., Ali, S., & Erdodi, L. A. (2019). Geographic variation and instrumentation artifacts:in search of confounds in performance validity assessment in adults with mild TBI. Advance online publication. *Psychological Injury and Law*, https://doi.org/10.1007/s12207-019-0935.

Lippa, S. M., Agbayani, K. A., Hawes, S., Jokic, E., & Caroselli, J. S. (2014). Effort in acute traumatic brain injury: Considering more than Pass/Fail. *Rehabilitation Psychology, 59*(3), 306–312.

Locke, D. E., Smigielski, J. S., Powell, M. R., & Stevens, S. R. (2008). Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *NeuroRehabilitation, 23*(3), 273–281.

Lu, P. H., Boone, K. B., Cozolino, L., & Mitchell, C. (2003). Effectiveness of the Rey-Osterrieth complex figure test and the Meyers and Meyers recognition trial in the detection of suspect effort. *The Clinical Neuropsychologist, 17*(3), 426–440. https://doi.org/10.1076/clin.17.3.426.18083.

Matthews, C. G., & Klove, H. (1964). *Instruction manual for the adult neuropsychology test battery* (p. 36). Madison: University of Wisconsin Medical School.

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Achieves of Clinical Neuropsychology, 27*(1), 10–22. https://doi.org/10.1093/arclin/acr084.

Mittenberg, W., Rotholc, A., Russell, E., & Heilbronner, R. (1996). Identification of malingered head injury on the Halstead-Reitan battery. *Archives of Clinical Neuropsychology, 11*, 271–281.

Newcombe, F. (1969). *Missile wounds of the brain*. London: Oxford University Press.

Odland, A. P., Lammy, A. B., Martin, P. K., Grote, C. L., & Mittenberg, W. (2015). Advanced administration and interpretation of multiple validity tests. *Psychological Injury and Law, 8*, 46–63.

Ord, J. S., Boettcher, A. C., Greve, K. J., & Bianchini, K. J. (2010). Detection of malingering in mild traumatic brain injury with the Conners' continuous performance test-II. *Journal of Clinical and Experimental Neuropsychology, 32*(4), 380–387.

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. Advance online publication. *The Clinical Neuropsychologist, 32*(6), 1039–1053. https://doi.org/10.1080/13854046.2017.1419507.

Prigatano, G. P., & Borgaro, S. R. (2003). Qualitative features of finger movement during the Halstead finger oscillation test following traumatic brain injury. *Journal of the International Neuropsychological Society, 9*(1), 128–133.

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

Rai, J. K., & Erdodi, L. A. (2019). The impact of criterion measures on the classification accuracy of TOMM-1. *Applied Neuropsychology: Adult*, https://doi.org/10.1080/23279095.2019.161.1613994.

Rai, J., An, K. Y., Charles, J., Ali, S., & Erdodi, L. A. (2019). Introducing a forced choice recognition trial to the Rey Complex Figure Test. *Psychology and Neuroscience*, https://doi.org/10.1037/pne0000175.

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. https://doi.org/10.1076/jcen.20.1.89.1488.

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*, 30–37. https://doi.org/10.1093/arclin/acs106.

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.

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology, 19*, 393–394.

Reitan, R. M. (1958). The validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8*, 271–276.

Reitan, R. M. (1969). *Manual for administration of neuropsychological test batteries for adults and children*. Indianapolis, IN.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Theory and interpretation*. Tucson: Neuropsychology Press.

Ruff, R. M., & Parker, S. B. (1993). Gender-and age-specific changes in motor speed and eye-hand coordination in adults: Normative values for the finger tapping and grooved pegboard tests. *Perceptual and Motor Skills, 76*(3), 1219–1230.

Schatz, P. (2011). *Encyclopedia of neuropsychology: Finger tapping test*. New York: Springer New York.

Schutte, C., Axelrod, B. N., & Montoya, E. (2015). Making sure neuropsychological data are meaningful: Use of performance validity testing in medicolegal and clinical contexts. *Psychological Injury and Law, 8*(2), 100–105.

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

Shimoyama, I., Yoshida, A., Yugeta, T., Saeki, N., Hayashi, F., Yoshizaki, H., & Shimizu, R. (2012). Rapid finger-tapping test and aging. *International Medical Journal, 19*(2), 138–140.

Shura, R. D., Miskey, H. M., Rowland, J. A., Yoash-Gatz, R. E., & Denning, J. H. (2016). Embedded performance validity measures

with postdeployment veterans: Cross-validation 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.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests*. New York: Oxford University Press.

Suhr, J. A. (2003). Neuropsychological impairment in fibromyalgia. Relation to depression, fatigue, and pain. *Journal of Psychosomatic Research, 55*, 321–329.

Suhr, J., & Spickard, B. (2012). Pain-related fear is associated with cognitive task avoidance: Exploration of the cogniphobia construct in a recurrent headache sample. *The Clinical Neuropsychologist, 26*(7), 1128–1141. https://doi.org/10.1080/13854046.2012.713121.

Sussman, Z. W., Peterson, R. L., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2019). Utility of Matrix Reasoning as an embedded performance validity indicator in pediatric mild traumatic brain injury. *Applied Neuropsychology: Child, 8*(1), 70–75. https://doi.org/10.1080/21622965.2017.1382359.

Sweet, J. J., Goldman, D. J., & Guidotti Breting, L. M. (2013). Traumatic brain injury: Guidance in a forensic context from outcome, dose response, and response bias research. *Behavioural Sciences and the Law, 31*, 756–778.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist, 16*(4), 495–505. https://doi.org/10.1076/clin.16.4.495.13909.

Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered 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.

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.

Wechsler, D. (2008). *Wechsler Adult Intelligence Test — Fourth Edition (WAIS-IV)*. San Antonio: Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale — Fourth Edition (WMS–IV)*. San Antonio: Pearson.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test 4*. Lutz: Psychological Assessment Resources, Inc.

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.