



# The Importance of Demographically Adjusted Cutoffs: Age and Education Bias in Raw Score Cutoffs Within the Trail Making Test

Christopher Abeare<sup>1</sup> · Alana Sabelli<sup>2</sup> · Brian Taylor<sup>2</sup> · Matthew Holcomb<sup>3</sup> · Claudiu Dumitrescu<sup>4</sup> · Ned Kirsch<sup>5</sup> · Laszlo Erdodi<sup>1</sup>

Received: 1 May 2019 / Accepted: 10 May 2019 / Published online: 8 June 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

This study was designed to develop validity cutoffs by utilizing demographically adjusted T-scores on the trail making test (TMT), with the goal of eliminating potential age and education-related biases associated with the use of raw score cutoffs. Failure to correct for the effect of age and education on TMT performance may lead to increased false positive errors for older adults and examinees with lower levels of education. Data were collected from an archival sample of 100 adult outpatients ( $M_{Age} = 38.8$ , 56% male;  $M_{Ed} = 13.7$ ) who were clinically referred for neuropsychological assessment at an academic medical center in the Midwestern USA after sustaining a traumatic brain injury (TBI). Performance validity was psychometrically determined using the Word Memory Test and two multivariate validity composites based on five embedded performance validity indicators. Cutoffs on the demographically corrected TMT T-scores had generally superior classification accuracy compared to the raw score cutoffs reported in the literature. As expected, the T-scores also eliminated age and education bias that was observed in the raw score cutoffs. Both T-score and raw score cutoffs were orthogonal to injury severity. Multivariate models of T-score based cutoff failed to improve classification accuracy over univariate T-score cutoffs. The present findings provide support for the use of demographically adjusted validity cutoffs within the TMT. They produced superior classification to raw score-based cutoffs, in addition to eliminating the bias against older adults and examinees with lower levels of education.

**Keywords** Trail making test · Performance validity · Demographic correction · Traumatic brain injury · Erdodi Index

The trail making test (TMT) is a commonly used neuropsychological test (Tombaugh, 2004) that is sensitive to brain dysfunction arising from a number of conditions, such as traumatic brain injury (Sanchez-Cubillo et al., 2009; Woods, Wyma, Herron, & Yund, 2015), schizophrenia (Sanchez-Cubillo et al., 2009), and dementia (Rasmusson, Zonderman, Kawas, & Resnick, 1998; Ashendorf et al., 2008). In addition to being an indicator of brain dysfunction and a measure of

executive functioning, attention, and visuospatial skills, cutoffs have been established on the TMT in order to function as an embedded validity indicator (EVI; Ruffolo, Guilmette, & Willis, 2000; Iverson, Lange, Green, & Frazen, 2002; O'Bryant, Hilsabeck, Fisher, & McCaffrey, 2003; Merten, Bossink, & Schmand, 2007; Powell, Locke, Smigielski, & McCrea, 2011; Busse & Whiteside, 2012; Ashendorf, Clark, & Sugarman, 2017). The use of the TMT as an EVI is consistent with the general trend of expanding performance validity assessment beyond traditional forced choice recognition memory paradigms and is supported by literature showing that processing speed measures can be useful in the detection of invalid performance (Ashendorf et al., 2017; Erdodi et al., 2017; Erdodi & Lichtenstein, 2017; Erdodi, Abeare, Lichtenstein, Erdodi, & Linnea, 2017).

Performance validity is the degree to which scores on cognitive tests reflect the test taker's actual ability (Boone, 2013; Larrabee, 2012). There are a number of factors that can lead to invalid performance, including inattention, lack of motivation to provide one's best performance, limited English proficiency (Erdodi, Jongasma, & Issa, 2017; Erdodi, Nussbaum, Sagar,

✉ Laszlo Erdodi  
lerdodi@gmail.com

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

<sup>2</sup> Brain-Cognition-Neuroscience Program, University of Windsor, Windsor, ON N9B 3P4, Canada

<sup>3</sup> Jefferson Neurobehavioral Group, New Orleans, LA, USA

<sup>4</sup> Canandgua Veteran's Administration Hospital, Rochester, NY, USA

<sup>5</sup> Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, MI, USA

Abeare, & Schwartz, 2017; Salazar, Lu, Wen, & Boone, 2007), and intentional suppression of performance (i.e., malingering). Administration of performance validity tests (PVTs) in all neuropsychological evaluations are considered best practice and are to be included as a basic standard of care (American Academy of Clinical Neuropsychology, 2007; Boone, 2009; Bush, Heilbronner, & Ruff, 2014; Chafetz et al., 2015; Heilbronner et al., 2009; Schutte, Axelrod, & Montoya, 2015).

PVTs can be stand-alone measures specifically designed to detect invalid performance. In many circumstances, administration of stand-alone PVTs is considered a requisite component of a complete neuropsychological evaluation. However, the use of stand-alone PVTs incurs increased time and financial costs, while providing only a snapshot of performance validity at a discrete moment. In forensic practice, this has led to an average use of 2.4 standalone PVTs in a battery, placing more time demands on clinicians and patients alike (Martin, Schroeder, & Odland, 2015).

For this and many other reasons, EVIs have been developed to supplement PVTs in a neuropsychological assessment. EVIs have the benefit of providing a more continuous measurement of performance validity across a test battery and require no additional resources (Lichtenstein et al., 2017). However, EVIs have been criticized on the grounds that they have higher rates of false positives due to factors such as credible cognitive impairment (Bigler, 2012; Bigler, 2015), limited English proficiency (Erdodi, Nussbaum, et al., 2017), or other factors known to influence test scores, such as age (Abeare, Messa, Zuccato, Merker, & Erdodi, 2018; Lichtenstein, Holcomb, & Erdodi, 2018). Although some of these issues are more easily addressed than others, there is a straightforward potential solution to address the influence of demographics: use EVIs based on demographically corrected scores.

By definition, EVIs based on demographically adjusted scores are less likely to be influenced by patient characteristics taken into account through stratified norming procedures. Controlling for key demographic factors is especially important with the TMT, as it is known to be influenced by age and education (Heaton, Miller, Taylor, & Grant, 2004; Tombaugh, 2004; Sanchez-Cubillo et al., 2009; Cavaco et al., 2013). Age has been demonstrated to be positively correlated with time to completion (Yuspeh, Drane, Huthwaite, & Klingler, 2000; MacNeill Horton & Roberts, 2001; Tombaugh, 2004; Hester, Kinsella, Ong, & McGregor, 2005; Perianez et al., 2007; Hamdan & Hamdan, 2009). Similarly, education has been shown to be negatively correlated with TMT completion times (MacNeill Horton & Roberts, 2001; Hamdan & Hamdan, 2009). However, the findings with regard to education have been less consistent: some studies report that only the TMT-B is influenced by education (Tombaugh, 2004; Hester et al., 2005; Hashimoto et al., 2006).

A number of existing EVIs already make use of demographically adjusted scores in order to avoid specific

demographic biases (Arnold et al., 2005; Erdodi et al., 2017; Erdodi, Kirsch, Sabelli, & Abeare, 2018). However, until recently, there were no such cutoffs established for the TMT. Ashendorf et al. (2017) appear to be the first to do so, correcting for age, based on data published by Mitrushina, Boone, Razani, and D'Elia (2005). Although this adjustment appears to improve diagnostic accuracy while correcting for the confounding effect of age, educational adjustments may be needed as well. The present study was designed to fill the gap in the current research literature by establishing cutoffs based on age *and* education corrected T-scores.

## Method

### Participants

A consecutive case series of 100 clinically referred patients were evaluated following a TBI at an academic medical center in the Midwestern USA. The majority 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). Most patients in the sample were classified as mild TBI (76%) by the assessing neuropsychologist based on commonly used guidelines (GCS > 13, LOC < 30 min, and PTA < 24 h). Data on litigation status were not available for the majority of the sample, given that all patients were referred by physicians. The present sample overlaps with previous publications from the same research group. The focus of the previous studies was different from the current study: the Forced Choice Recognition trial of the California Verbal Learning Test–Second Edition (CVLT-II; Erdodi, Abeare, Medoff et al., 2018), the Grooved Pegboard Test (GPB; Erdodi, Kirsch, Sabelli, & Abeare, 2018), and the Conner's Continuous Performance Test–Second Edition (CPT-II; Erdodi, Roth, Kirsch, Lajiness-O'Neill, & Medoff, 2014).

### Materials

The core battery of neuropsychological tests administered to all patients included the Wechsler Adult Intelligence (WAIS-IV; Wechsler, 2008), the Wechsler Memory Scale–Fourth Edition (WMS-IV; Wechsler, 2009), the CVLT-II (Delis, Kramer, Kaplan, & Ober, 2000), CPT-II (Conners, 2004), the trail making test (TMT A & B; Retain, 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), 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 Finger Tapping Test and GPB, using demographically adjusted T-scores based on norms by Heaton et al. (2004). The computer-administered version of the Word Memory Test (Green, 2003) was administered as the primary PVT.

A variety of methods have been utilized to detect invalid responding using the TMT. Time to completion of TMT-A, TMT-B, and combined trials (TMT A + B) have all been previously identified as effective measures to determine performance validity (O’Bryant et al., 2003; Powell et al., 2011; Busse & Whiteside, 2012; Shura, Miskey, Rowland, Yoash-Gatz, & Denning, 2016). Additionally, measures of the difference between the scores (Trails B–A; Ruffolo et al., 2000; Yuspeh et al., 2000) and ratio (Trails B/A; Corrigan & Hinkeldey, 1987; Ruffolo et al., 2000) have also gained some support as methods to identifying non-credible performance.

In order to continuously monitor performance validity throughout the assessment (Boone, 2009; Bush et al., 2014; Chafetz et al., 2015; Schutte et al., 2015) and to provide a comprehensive evaluation of performance validity across the cognitive domains (Cottingham, Victor, Boone, Ziegler, & Zeller, 2014; Webber, Critchfield, & Soble, 2018), two

validity composites were created (“Erdodi Index Five” or EI-5), following the methodology outlined by Erdodi (2019). One was based on PVTs embedded within measures of visuomotor processing speed and was labeled “EI-5<sub>VM</sub>.” The other was based on PVTs based on forced choice recognition paradigms and was labeled “EI-5<sub>FCR</sub>.”

EI-5 components were recoded onto a four-point ordinal scale, where zero was defined as an incontrovertible *Pass* and three as an incontrovertible *Fail*. An EI-5 value of one signifies failing the most liberal (i.e., high sensitivity, low specificity) cutoff, which is typically associated with a base rate of failure (BR<sub>Fail</sub>) around 25% (An et al., 2019; Erdodi et al., 2018; Erdodi et al., 2016; 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<sub>Fail</sub> of 10 and 5, respectively (Table 1). 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.

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

**Table 1** The components of the EI-5s and base rates of failure at given cutoffs

*The Components of the EI-5s and Base Rates of Failure at Given Cutoffs*

| EI-5 <sub>VM</sub>    | EI-5 Value |        |         |             | EI-5 <sub>FCR</sub>    | EI-5 Value |       |       |             |
|-----------------------|------------|--------|---------|-------------|------------------------|------------|-------|-------|-------------|
|                       | 0          | 1      | 2       | 3           |                        | 0          | 1     | 2     | 3           |
| <i>Components</i>     | Pass       | Fail   | FAIL    | <b>FAIL</b> | <i>Components</i>      | Pass       | Fail  | FAIL  | <b>FAIL</b> |
| CD <sub>WAIS-IV</sub> | ≥6         | 5      | 4       | ≤3          | FCR <sub>CVLT-II</sub> | 16         | 14-15 | 13    | ≤12         |
| <i>Base Rate</i>      | 75%        | 12%    | 6%      | 7%          | <i>Base Rate</i>       | 74.0%      | 15.0% | 5.0%  | 6.0%        |
| COM <sub>CPT-II</sub> | ≥60        | 61-67  | 68-72   | ≤73         | LM <sub>WMS-IV</sub>   | ≥21        | 19-20 | 18    | ≤17         |
| <i>Base Rate</i>      | 77%        | 13%    | 6%      | 4%          | <i>Base Rate</i>       | 81.0       | 11.0% | 3.0%  | 5.0%        |
| GPB (DH)              | ≥30        | 23-29  | 18-22   | ≤17         | RH <sub>CVLT-II</sub>  | ≥11        | 9-10  | 6-8   | ≤5          |
| <i>Base Rate</i>      | 80%        | 11%    | 5%      | 4%          | <i>Base Rate</i>       | 80.0       | 10.0% | 7.0%  | 3.0%        |
| OMI <sub>CPT-II</sub> | ≥65        | 66-150 | 151-233 | ≤234        | VPA <sub>WMS-IV</sub>  | ≥35        | 29-34 | 27-28 | ≤26         |
| <i>Base Rate</i>      | 73%        | 17%    | 5%      | 5%          | <i>Base Rate</i>       | 74.0       | 17.0% | 4.0%  | 5.0%        |
| SS <sub>WAIS-IV</sub> | ≥7         | 5-6    | 3-4     | ≤2          | VR <sub>WMS-IV</sub>   | ≥5         | 3-4   | 2     | ≤1          |
| <i>Base Rate</i>      | 75%        | 16%    | 7%      | 5%          | <i>Base Rate</i>       | 67.0       | 26.0% | 3.0%  | 4.0%        |

*Note.* EI-5<sub>VM</sub>: Erdodi Index Five–Visuomotor; CD<sub>WAIS-IV</sub>: age-corrected scaled score on the coding subtest of the Wechsler Adult Intelligence Scale–Fourth Edition (Erdodi, Abeare et al., 2017; Erdodi & Lichtenstein, 2017; Etherton, Bianchini, Heinly, & Greve, 2006; Kim et al., 2010; Trueblood, 1994); COM<sub>CPT-II</sub>: Conners’ Continuous Performance Test – Second Edition Omissions T-scores (Erdodi, Roth et al., 2014; Erdodi & Roth, 2017; Lange et al., 2013; Marshall et al., 2010); GPB (DH): demographically corrected T-scores (Heaton et al., 2004) on the Grooved Pegboard Test, using the dominant hand (Erdodi, Seke et al., 2017) OMI<sub>CPT-II</sub>: Conners’ Continuous Performance Test, 2nd edition Omissions T-scores (Erdodi, Roth et al., 2014; Erdodi & Roth, 2017; Lange et al., 2013; Marshall et al., 2010; Ord, Boettcher, Greve, & Bianchini, 2010); SS<sub>WAIS-IV</sub>: age-corrected scaled score on the Symbol Search subtest of the Wechsler Adult Intelligence Scale–Fourth Edition (Erdodi, Abeare et al., 2017; Erdodi & Lichtenstein, 2017; Etherton et al., 2006; Trueblood, 1994); EI-5<sub>FCR</sub>: Erdodi Index Five–Forced Choice Recognition; FCR<sub>CVLT-II</sub>: Forced Choice Recognition trial of the California Verbal Learning Test–Second Edition (raw score; Erdodi et al., 2017; Schwartz et al., 2016); LM<sub>WMS-IV</sub>: Logical Memory recognition trial from the Wechsler Memory Scales–Fourth Edition (raw score; Bortnik et al., 2010; Pearson, 2009); RH<sub>CVLT-II</sub>: Yes/No recognition hits (raw score; Greve, Curtis, Bianchini, & Ord, 2009; Wolfe et al., 2010); VPA<sub>WMS-IV</sub>: Verbal Paired Associates recognition trial (raw score; Pearson, 2009); VR<sub>WMS-IV</sub>: Visual Reproduction recognition trial (raw score; Pearson, 2009)

**Table 2** Frequency, cumulative frequency and classification range for the first ten levels of the EI-5s*Frequency, Cumulative Frequency and Classification Range for the First Ten Levels of the EI-5s*

| EI-5 | EI-5 <sub>VM</sub> |              | EI-5 <sub>FCR</sub> |              | Classification |         |
|------|--------------------|--------------|---------------------|--------------|----------------|---------|
|      | <i>f</i>           | Cumulative % | <i>f</i>            | Cumulative % | By Row         | OVERALL |
| 0    | 46                 | 46           | 45                  | 45           | PASS           | PASS    |
| 1    | 15                 | 61           | 14                  | 59           | Pass           |         |
| 2    | 12                 | 73           | 9                   | 68           | Borderline     | FAIL    |
| 3    | 8                  | 81           | 9                   | 77           | Borderline     |         |
| 4    | 4                  | 85           | 7                   | 84           | Fail           |         |
| 5    | 2                  | 87           | 6                   | 90           | Fail           |         |
| 6    | 3                  | 90           | 1                   | 91           | FAIL           |         |
| 7    | 2                  | 92           | 1                   | 92           | FAIL           |         |
| 8    | 2                  | 94           | 2                   | 94           | FAIL           |         |
| 9    | 2                  | 96           | 2                   | 96           | FAIL           |         |

Note. EI-5<sub>VM</sub>: Erdodi Index Five–Visuomotor; EI-5<sub>FCR</sub>: Erdodi Index Five–Forced Choice Recognition

(i.e., all five constituent PVTs were failed at the most conservative cutoff). An EI-5 ≤ 1 is classified as an overall *Pass*, as it contains at most one failure at the most liberal cutoff. The interpretation of EI-5 values 2 and 3 is more problematic, as they can either indicate a single failure at a more conservative, 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 et al., 2018; Erdodi, Hurtubise, et al., 2018; Erdodi, Seke, Shahein, et al., 2017; Erdodi et al., 2018; Proto et al., 2014). Therefore, it is labeled *Borderline* and excluded from analyses requiring a dichotomous (Pass/Fail) criterion variable. 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 approximately 20% BR<sub>Fail</sub> (Table 2).

Due to the fact that the EI-model is a relatively new approach to multivariate performance validity assessment, it was validated against the WMT, a well-established free-standing PVT (Green, Iverson, & Allen, 1999; Iverson, Green, & Gervais, 1999; Tan, Slick, Strauss, & Hultsch, 2002). The EI-5<sub>FCR</sub> produced a slightly higher (.88) overall classification accuracy than the EI-5<sub>VM</sub> (.84), driven by superior sensitivity (.75 vs. .64). However, both versions of the EI-5 were highly specific (.96 and .98, respectively) to failure on the WMT (Table 3). Significant differences emerged on several TMT scores as a function of EI-5 level (Pass, Borderline, Fail). The contrasts

were associated with notably larger main effects ( $\eta^2$ : .33-.44, extremely large) for the EI-5<sub>VM</sub> (Table 4) as compared to the EI-5<sub>FCR</sub> ( $\eta^2$ : .12-.20, large effects; Table 5). All post hoc contrasts between Pass and Fail were significant, although effect sizes were more pronounced for the EI-5<sub>VM</sub> ( $d$ : 1.57-1.73, very large) than the EI-5<sub>FCR</sub> ( $d$ : 0.74-0.97, large). Likewise, all post hoc contrasts between Pass and Borderline were significant. Again, larger effects were observed for the EI-5<sub>VM</sub> ( $d$ : 0.75-1.15, large) than the EI-5<sub>FCR</sub> ( $d$ : 0.66-0.91, medium-large). This pattern of finding reinforces the notion that a third category (“indeterminate range”) is a legitimate outcome in performance validity assessment in addition to the traditional Pass/Fail dichotomy (Erdodi, 2019).

**Table 3** Classification accuracy of the EI-5s against the WMT

| EI version          | Classification Accuracy | Criterion PVT WMT |
|---------------------|-------------------------|-------------------|
| EI-5 <sub>VM</sub>  | AUC                     | .84               |
|                     | 95% CI for AUC          | .76-.93           |
|                     | Sensitivity             | .64               |
|                     | Specificity             | .98               |
| EI-5 <sub>FCR</sub> | AUC                     | .88               |
|                     | 95% CI for AUC          | .81-.96           |
|                     | Sensitivity             | .75               |
|                     | Specificity             | .96               |

PVT performance validity test, EI-5<sub>VM</sub> Erdodi Index Five–Visuomotor (*Fail* defined as ≥ 4; Erdodi, 2019), EI-5<sub>FCR</sub> Erdodi Index Five–Forced Choice Recognition (*Fail* defined as ≥ 4; Erdodi, 2019), AUC area under the curve (overall classification accuracy), WMT Word Memory Test (standard cutoffs)

**Table 4** Results of one-way ANOVAs on TMT scores across EI-5<sub>VM</sub> classification ranges

| TMT trial    | Score |    | EI-5 <sub>VM</sub>    |                      |                       | F    | P      | $\eta^2_p$ | Significant post hoc contrasts | d    |
|--------------|-------|----|-----------------------|----------------------|-----------------------|------|--------|------------|--------------------------------|------|
|              |       |    | 0–1<br>n = 61<br>PASS | 2–3<br>n = 20<br>BOR | ≥ 4<br>n = 19<br>FAIL |      |        |            |                                |      |
| Trails A     | T     | M  | 48.7                  | 40.7                 | 28.7                  | 23.9 | < .001 | .33        | PASS vs. BOR                   | .85  |
|              |       | SD | 11.0                  | 7.7                  | 14.4                  |      |        |            |                                |      |
| Trails B     | T     | M  | 49.8                  | 36.7                 | 30.3                  | 24.5 | < .001 | .34        | PASS vs. BOR                   | 1.15 |
|              |       | SD | 11.1                  | 11.7                 | 13.2                  |      |        |            |                                |      |
| Trails A     | Raw   | M  | 27.6                  | 35.9                 | 59.0                  | 35.0 | < .001 | .42        | PASS vs. BOR                   | .75  |
|              |       | SD | 10.8                  | 11.2                 | 24.1                  |      |        |            |                                |      |
| Trails B     | Raw   | M  | 63.5                  | 107.9                | 165.6                 | 33.1 | < .001 | .41        | PASS vs. BOR                   | 1.03 |
|              |       | SD | 24.4                  | 55.6                 | 87.0                  |      |        |            |                                |      |
| Trails A + B | Raw   | M  | 91.1                  | 143.8                | 224.6                 | 38.5 | < .001 | .44        | PASS vs. BOR                   | 1.06 |
|              |       | SD | 32.7                  | 62.5                 | 104.2                 |      |        |            |                                |      |
|              |       |    |                       |                      |                       |      |        |            | BOR vs. FAIL                   | .94  |

TMT trail making test; post hoc pairwise contrasts were computed using the *least significant difference* method, BOR borderline,  $\eta^2_p$  partial eta squared, EI-5<sub>VM</sub> Erdodi Index Five–Visuomotor, T demographically adjusted score using the norms by Heaton et al. (2004)

## Procedure

Patients were assessed in two half-day (4-hour) appointments in an outpatient setting. Testing was conducted by Master's level psychologists with specialty training in psychometric testing. De-identified data were collected retrospectively. The project was approved by the relevant institutional boards).

Significant differences emerged on several TMT scores as a function of EI-5 level (Pass, Borderline, Fail). The contrasts

were associated with notably larger main effects ( $\eta^2_p$ : .33-.44, extremely large) for the EI-5VM as compared to the EI-5FCR ( $\eta^2_p$ : .12-.20, large effects; Table 5). All post hoc contrasts between Pass and Fail were significant, although effect sizes were more pronounced for the EI-5VM (d: 1.57-1.73, very large) than the EI-5FCR (d: 0.74-0.97, large). Likewise, all post hoc contrasts between Pass and Borderline were significant. Again, larger effects were observed for the EI-5VM (d: 0.75-1.15, large) than the EI-5FCR (d: 0.66-

**Table 5** Results of one-way ANOVAs on TMT scores across EI-5<sub>FCR</sub> classification ranges

| TMT trial    | Score |    | EI-5 <sub>FCR</sub>   |                      |                       | F    | p      | $\eta^2_p$ | Significant post hoc contrasts | d   |
|--------------|-------|----|-----------------------|----------------------|-----------------------|------|--------|------------|--------------------------------|-----|
|              |       |    | 0–1<br>n = 59<br>PASS | 2–3<br>n = 18<br>BOR | ≥ 4<br>n = 23<br>FAIL |      |        |            |                                |     |
| Trails A     | T     | M  | 47.5                  | 39.2                 | 35.9                  | 8.11 | < .001 | .14        | PASS vs. BOR                   | .66 |
|              |       | SD | 11.5                  | 13.4                 | 14.8                  |      |        |            |                                |     |
| Trails B     | T     | M  | 47.5                  | 38.0                 | 37.4                  | 6.43 | .002   | .12        | PASS vs. BOR                   | .69 |
|              |       | SD | 12.5                  | 14.8                 | 14.7                  |      |        |            |                                |     |
| Trails A     | Raw   | M  | 28.7                  | 40.3                 | 47.9                  | 11.7 | < .001 | .19        | PASS vs. BOR                   | .76 |
|              |       | SD | 11.4                  | 18.2                 | 25.5                  |      |        |            |                                |     |
| Trails A     | Raw   | M  | 70.3                  | 109.7                | 133.0                 | 11.1 | < .001 | .19        | PASS vs. BOR                   | .89 |
|              |       | SD | 40.1                  | 48.3                 | 91.3                  |      |        |            |                                |     |
| Trails A + B | Raw   | M  | 99.0                  | 150.0                | 180.9                 | 12.5 | < .001 | .20        | PASS vs. BOR                   | .91 |
|              |       | SD | 48.4                  | 62.7                 | 111.7                 |      |        |            |                                |     |

TMT trail making test; post hoc pairwise contrasts were computed using the *least significant difference* method, BOR borderline,  $\eta^2_p$  partial eta squared, EI-5<sub>FCR</sub> Erdodi Index Five–Forced Choice Recognition, T demographically adjusted score using the norms by Heaton et al. (2004)

**Table 6** Classification accuracy of various TMT validity cutoffs

| TMT Trial | Score | Cutoff | BR <sub>Fail</sub> | Criterion PVT |      |                             |      |                              |      |
|-----------|-------|--------|--------------------|---------------|------|-----------------------------|------|------------------------------|------|
|           |       |        |                    | WMT<br>39.0   |      | EI-5 <sub>VM</sub><br>23.8% |      | EI-5 <sub>FCR</sub><br>28.1% |      |
|           |       |        |                    | SENS          | SPEC | SENS                        | SPEC | SENS                         | SPEC |
| Trails A  | T     | ≤ 35   | 23.0               | .36           | .85  | .74                         | .92  | .48                          | .88  |
|           |       | ≤ 34   | 21.0               | .36           | .89  | .74                         | .93  | .48                          | .92  |
|           |       | ≤ 33   | 19.0               | .33           | .90  | .68                         | .95  | .43                          | .93  |
| Trails B  | T     | ≤ 35   | 28.0               | .49           | .85  | .68                         | .90  | .48                          | .86  |
|           |       | ≤ 34   | 23.0               | .38           | .87  | .63                         | .93  | .39                          | .88  |
|           |       | ≤ 33   | 21.0               | .38           | .90  | .63                         | .95  | .39                          | .90  |
| Trails A  | Raw   | ≥ 63"  | 10.0               | .21           | .97  | .42                         | .98  | .30                          | .98  |
| Trails B  | Raw   | ≥ 120" | 22.0               | .36           | .87  | .63                         | .93  | .39                          | .90  |
|           |       | ≥ 200" | 6.0                | .10           | .97  | .26                         | 1.00 | .17                          | .98  |
|           |       | ≥ 137" | 31.0               | .54           | .84  | .82                         | .90  | .57                          | .85  |
| A + B     | Raw   | ≥ 170" | 19.0               | .36           | .92  | .63                         | .95  | .39                          | .93  |
| Trails    | Raw   | < 1.50 | 3.0                | .00           | .95  | .00                         | .95  | .00                          | .95  |
| B/A       |       | < 1.65 | 6.0                | .03           | .92  | .05                         | .92  | .09                          | .93  |
|           |       | < 1.70 | 10.0               | .10           | .90  | .11                         | .87  | .17                          | .91  |
| B–A       | T     | ≥ 10   | 18.0               | .21           | .84  | .32                         | .80  | .35                          | .85  |
|           |       | ≥ 12   | 12.0               | .10           | .87  | .16                         | .85  | .17                          | .88  |
|           |       | ≥ 15   | 6.0                | .08           | .92  | .11                         | .93  | .04                          | .93  |
| T3-LIB    | T     | ≥ 2    | 21.0               | .36           | .89  | .74                         | .93  | .48                          | .92  |
| T3-CON    | T     | ≥ 2    | 16.0               | .28           | .92  | .58                         | .97  | .35                          | .95  |

PVT performance validity test, WMT Word Memory Test (standard cutoffs), EI-5<sub>VM</sub> Erdodi Index Five–Visuomotor (*Fail* defined as ≥ 4; Erdodi, 2019), EI-5<sub>FCR</sub> Erdodi Index Five–Forced Choice Recognition (*Fail* defined as ≥ 4; Erdodi, 2017), TMT trail making test, T3-LIB number of failures at the following cutoffs: TMT-A T ≤ 35; TMT-B T ≤ 35; TMT B–A T ≥ 10, T3-CON number of failures at the following cutoffs: TMT-A T ≤ 33; TMT-B T ≤ 33; TMT B–A T ≥ 15, T demographically adjusted T-score (Heaton et al., 2004), SENS sensitivity, SPEC specificity

0.91, medium-large). This pattern of finding reinforces the notion that a third category (“indeterminate range”) is a legitimate outcome in performance validity assessment in addition to the traditional Pass/Fail dichotomy (Erdodi, 2019).

## Data Analysis

Descriptive statistics (mean, standard deviation, range, BR<sub>Fail</sub>) were reported for key variables. Overall classification accuracy (AUC) and the corresponding 95% confidence intervals were computed using SPSS version 23.0. AUC values in the .70–.79 range were classified as *acceptable*; a value in the .80–.89 range was classified as *excellent*, whereas an AUC ≥ .90 was 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. One-way ANOVAs and *t* tests were computed to compare means, as appropriate).

## Results

### Raw Score Cutoffs

The raw score-based EVIs for TMT A and B performed reasonably well overall against the three criterion PVTs, with sensitivity ranging from .10 to .63 and specificity ranging from .87 to .98 (Table 6). Classification accuracy was best against the EI-5<sub>VM</sub>, as expected due to modality specificity (Erdodi, 2019). The TMT A + B had the best classification accuracy of the raw score-based EVIs, with sensitivity ranging from .36 to .82 and specificity ranging from .84 to .95. Again, classification accuracy was best against the EI-5<sub>VM</sub>. The TMT B/A had unacceptably poor classification accuracy and was therefore excluded from subsequent analyses.

**Table 7** Base rate of failure (%) as a function of injury severity across various PVTs and cutoffs

| PVT                 | Score    | Cutoff   | Injury severity       |                      | RR   | $\chi^2$ | <i>p</i> |
|---------------------|----------|----------|-----------------------|----------------------|------|----------|----------|
|                     |          |          | <i>n</i> = 74<br>Mild | <i>n</i> = 26<br>M/S |      |          |          |
| WMT                 |          | Standard | 46.1                  | 16.7                 | 2.76 | 6.62     | .008     |
| EI-5 <sub>VM</sub>  |          | ≥ 4      | 29.3                  | 9.1                  | 3.22 | 3.60     | .049     |
| EI-5 <sub>FCR</sub> |          | ≥ 4      | 32.3                  | 15.0                 | 2.15 | 2.23     | .111     |
| Trails A            | Raw      | ≥ 63"    | 10.5                  | 8.3                  | 1.27 | 0.97     | .553     |
| Trails A            | <i>t</i> | ≤ 33     | 18.4                  | 20.8                 | 0.88 | 0.07     | .501     |
| Trails B            | Raw      | ≥ 200"   | 6.6                   | 4.2                  | 1.57 | 0.19     | .555     |
|                     |          | ≥ 120"   | 22.4                  | 20.8                 | 1.08 | 0.03     | .560     |
| Trails B            | <i>t</i> | ≤ 33     | 23.7                  | 12.5                 | 1.90 | 1.38     | .190     |
| Trails              | Raw      | ≥ 137"   | 31.6                  | 29.2                 | 1.08 | 0.05     | .519     |
| A + B               |          | ≥ 170"   | 19.7                  | 16.7                 | 1.18 | 0.11     | .499     |

WMT Green’s Word Memory Test, PVT performance validity test, EI-5<sub>VM</sub> Erdodi Index Five–Visuomotor (Erdodi, 2017), EI-5<sub>FCR</sub> Erdodi Index Five–Forced Choice Recognition (Erdodi, 2017), M/S moderate-to-severe, RR risk ratio

**T-Score Cutoffs**

Classification accuracy for the TMT A and B demographically corrected T-scores was generally better than raw score cutoffs and varied depending on the criterion PVT. TMT A T-scores were slightly less sensitive to failure on the WMT (.33 to .36) than TMT B T-scores (.38 to .49), but were slightly more sensitive to failure on the two EI-5s (.43 to .74 and .39 to .68, respectively). Specificity values on TMT A and B T-scores were

comparable across criterion PVTs. Cutoffs of  $T \leq 33$  on both TMT A and B optimized sensitivity (.33 and .38, respectively) with strong specificity ( $\geq 90$ ) against the WMT. The TMT B–A had fairly poor sensitivity (.10 to .17) at acceptable specificity values (.85 to .88).

Lastly, the three T-score-based EVIs were combined into a multivariate liberal (T3-LIB) composite in which the number of failures on liberal cutoffs was summed and a conservative multivariate composite (T3-CON) in which the number of failures of conservative cutoffs was summed, using an overall cutoff of  $\geq 2$  failures. The multivariate combination had comparable classification accuracy to the univariate cutoffs, with sensitivity ranging from .28 to .74 and specificity ranging from .89 to .97 (Table 7).

**Effects of Age and Education**

As expected, EVIs based on raw scores were influenced by age (*d*: .47–.84 medium-large). However, there was no effect of age on the T-score cutoffs (see Table 8). Similarly, there was no effect of education on TMT A raw or T-scores. However, there was an effect of education on TMT B raw scores and TMT A + B raw scores (*d*: .39–.52, medium). There was no effect of education on TMT A or B T-scores (see Table 9).

The effects of age and education were also examined for the criterion PVTs. There was a medium effect of age (*d* = .42) on the WMT and EI-5<sub>FCR</sub> (*d* = .45), but there was no effect on the EI-5<sub>VM</sub>. There were small to medium, but statistically non-significant, effects of education on the WMT (*d* = .38) and EI-5<sub>VM</sub> (*d* = .44) and a medium effect on the EI-5<sub>FCR</sub> (*d* = .63).

**Table 8** The relationship between failing TMT validity cutoffs and age of the examinee

| TMT Trial    | Score  | Cutoff | Outcome | <i>n</i> | Age (years) |           | <i>t</i> | <i>p</i> | <i>d</i> |
|--------------|--------|--------|---------|----------|-------------|-----------|----------|----------|----------|
|              |        |        |         |          | <i>M</i>    | <i>SD</i> |          |          |          |
| Trails A     | Raw    | < 63"  | Pass    | 90       | 37.5        | 13.7      | 2.34     | .021     | .84      |
|              |        | ≥ 63"  | Fail    | 10       | 48.1        | 11.5      |          |          |          |
|              | T      | > 33   | Pass    | 81       | 38.3        | 14.1      | 0.47     | .638     | –        |
|              |        | ≤ 33   | Fail    | 19       | 40.0        | 13.0      |          |          |          |
| Trails B     | Raw    | < 120" | Pass    | 78       | 37.2        | 14.0      | 1.98     | .050     | .49      |
|              |        | ≥ 120" | Fail    | 22       | 43.7        | 12.4      |          |          |          |
|              |        | < 200" | Pass    | 94       | 38.1        | 13.8      |          |          |          |
|              | ≥ 200" | Fail   | 6       | 46.8     | 13.7        |           |          |          |          |
|              | T      | > 33   | Pass    | 79       | 37.9        | 14.2      | 0.95     | .345     | –        |
|              |        | ≤ 33   | Fail    | 21       | 41.1        | 12.4      |          |          |          |
| Trails A + B | Raw    | < 137" | Pass    | 69       | 36.3        | 13.8      | 2.49     | .014     | .55      |
|              |        | ≥ 137" | Fail    | 31       | 43.6        | 12.7      |          |          |          |
|              | T      | < 170" | Pass    | 81       | 37.4        | 13.8      | 1.80     | .075     | .47      |
|              |        | ≥ 170" | Fail    | 19       | 43.7        | 13.2      |          |          |          |

Note. TMT: Trail Making Test; T: Demographically adjusted T-score (Heaton et al., 2004)

**Table 9** The Relationship between failing TMT validity cutoffs and the examinee’s level of education

| TMT trial    | Score | Cutoff | Outcome | n  | Education |     | T    | p    | d   |
|--------------|-------|--------|---------|----|-----------|-----|------|------|-----|
|              |       |        |         |    | M         | SD  |      |      |     |
| Trails A     | Raw   | < 63"  | Pass    | 90 | 13.7      | 2.4 | 0.83 | .412 | –   |
|              |       | ≥ 63"  | Fail    | 10 | 13.0      | 3.0 |      |      |     |
|              | T     | > 33   | Pass    | 81 | 13.8      | 2.4 | 1.52 | .131 | –   |
|              |       | ≤ 33   | Fail    | 19 | 12.8      | 2.4 |      |      |     |
| Trails B     | Raw   | < 120" | Pass    | 78 | 13.8      | 2.3 | 1.73 | .088 | .39 |
|              |       | ≥ 120" | Fail    | 22 | 12.8      | 2.8 |      |      |     |
|              | T     | < 200" | Pass    | 94 | 13.6      | 2.4 | 0.11 | .911 | –   |
|              |       | ≥ 200" | Fail    | 6  | 13.5      | 3.2 |      |      |     |
|              |       | > 33   | Pass    | 79 | 13.8      | 2.3 |      |      |     |
|              |       | ≤ 33   | Fail    | 21 | 13.0      | 2.9 |      |      |     |
| Trails A + B | Raw   | < 137" | Pass    | 69 | 14.0      | 2.3 | 2.42 | .017 | .52 |
|              |       | ≥ 137" | Fail    | 31 | 12.7      | 2.7 |      |      |     |
|              | T     | < 170" | Pass    | 81 | 13.8      | 2.4 | 1.95 | .054 | .47 |
|              |       | ≥ 170" | Fail    | 19 | 12.6      | 2.7 |      |      |     |

TMT trail making test, T demographically adjusted T-score (Heaton et al., 2004)

**Effects of Injury Severity**

Consistent with the literature, there was a reverse dose-response relationship (Hill, 1965; Carone, 2008; Erdodi & Rai, 2017; Green, Flaro, & Courtney, 2009; Green et al., 1999) between injury severity and the BR<sub>Fail</sub> on the WMT, with mTBI patients failing at a higher rate (46.1%) than those with moderate to severe TBIs (16.7%). The EI-5<sub>VM</sub> also had a higher BR<sub>Fail</sub> in the mTBI group (29.3%) than the moderate to severe group (9.1%). The EI-5<sub>FCR</sub> showed the same pattern, with the mTBI group more than twice as likely to fail, but this

difference was not statistically significant. It should, however, be noted that the Chi-square is sensitive to sample size, and in this case, it was underpowered. The BR<sub>Fail</sub> on the TMT did not vary as a function of TBI severity. (Table 10)

**Discussion**

EVI based on demographically corrected T-scores on the TMT were developed and compared to previously published raw score cutoffs in an attempt to reduce potential age and

**Table 10** The relationship between failing criterion PVTs and the examinee’s age and level of education

| PVT                 | Scale           | Cutoff   | Outcome | n  | Age (years) |      | t    | p    | d   |
|---------------------|-----------------|----------|---------|----|-------------|------|------|------|-----|
|                     |                 |          |         |    | M           | SD   |      |      |     |
| WMT                 | Raw             | Standard | Pass    | 61 | 36.4        | 14.3 | 2.01 | .047 | .42 |
|                     |                 |          | Fail    | 39 | 42.0        | 12.5 |      |      |     |
| EI-5 <sub>VM</sub>  | AC <sup>a</sup> | ≤ 1      | Pass    | 61 | 36.9        | 15.1 | 1.30 | .199 | –   |
|                     |                 | ≥ 4      | Fail    | 19 | 41.8        | 12.2 |      |      |     |
| EI-5 <sub>FCR</sub> | Raw             | ≤ 1      | Pass    | 59 | 36.4        | 14.9 | 1.78 | .080 | .45 |
|                     |                 | ≥ 4      | Fail    | 23 | 42.5        | 11.5 |      |      |     |
| Education           |                 |          |         |    |             |      |      |      |     |
| WMT                 | Raw             | Standard | Pass    | 61 | 14.0        | 2.6  | 1.75 | .083 | .38 |
|                     |                 |          | Fail    | 39 | 13.1        | 2.1  |      |      |     |
| EI-5 <sub>VM</sub>  | AC <sup>a</sup> | ≤ 1      | Pass    | 61 | 13.9        | 2.5  | 1.73 | .087 | .44 |
|                     |                 | ≥ 4      | Fail    | 19 | 12.8        | 2.5  |      |      |     |
| EI-5 <sub>FCR</sub> | Raw             | ≤ 1      | Pass    | 59 | 14.0        | 2.6  | 2.31 | .024 | .63 |
|                     |                 | ≥ 4      | Fail    | 23 | 12.6        | 1.8  |      |      |     |

WMT Word Memory Test, EI-5<sub>VM</sub> Erdodi Index Five–Visuomotor, EI-5<sub>FCR</sub> Erdodi Index Five–Forced Choice Recognition (Fail defined as ≥ 4; Erdodi, 2017)

<sup>a</sup> Although the EI-5<sub>VM</sub> scale itself is not demographically adjusted, all of its components are age-corrected



educational bias in the assessment of performance validity. Raw score cutoffs were found to be influenced by age and education, whereas the newly developed cutoffs for TMT A and TMT B demographically corrected T-scores successfully eliminated potential age and education biases. In addition, the T-score cutoffs had generally better classification accuracy than the raw-score cutoffs, with a better balance between sensitivity and specificity.

The so-called “derivative” EVIs (e.g., TMT A + B, B–A, B/A) had mixed support. The TMT B–A T-score difference and B/A raw score ratio had poor sensitivity at cutoffs with acceptable specificity. TMT A + B raw score had strong classification accuracy in this sample, but it was confounded by age and education, consistent with other raw-score cutoffs. Consequently, these EVIs have limited utility in a general clinical sample as they have an inflated false positive rate for older adults and examinees with lower levels of education.

The multivariate combinations, T3-LIB and T3-CON, had comparable classification accuracy compared to the univariate cutoffs. This is somewhat surprising, as multivariate combination of EVIs often results in improved classification accuracy (Boone, 2009; Erdodi et al., 2014; Rai, An, Charles, Ali, & Erdodi, 2019; Tyson et al., 2018). The reason for the lack of incremental utility in the multivariate combinations in this sample is unclear.

The criterion PVTs showed the expected reverse dose-response relationship with injury severity. The mTBI group had a 46.1%  $BR_{Fail}$  on the WMT compared to only 16.7%  $BR_{Fail}$  in the moderate to severe group ( $RR = 2.76$ ). The same pattern was found for the EI-5<sub>VM</sub> ( $BR_{Fail} = 29.3\%$  and  $9.1\%$ , respectively;  $RR = 3.22$ ) and the EI-5<sub>FCR</sub> ( $BR_{Fail} = 32.3\%$  and  $15.0\%$ ;  $RR = 2.15$ ), although the Chi-square for the EI-5<sub>FCR</sub> was non-significant ( $p = .111$ ). However, consistent with Iverson et al. (2002), the reverse dose-response relationship was not observed for any of the TMT-based EVIs. In fact, there were no reliable differences between severity groups on any of these EVIs. One can only speculate about possible reasons behind these negative findings and their clinical interpretation. The propensity to exaggerated deficits in patients with mTBI in the post-acute stage of recovery may have been canceled out by genuine persisting deficits in patients with moderate-to-severe TBI. It should be noted that the lack of difference in  $BR_{Fail}$  between these two levels of injury severity itself provides circumstantial evidence for non-credible performance in the mTBI sample, as the normative outcome is full return to baseline (Boone, 2013; McCrea, 2008). Therefore, a relative deficit in patients with moderate-to-severe TBI would be expected on such measures known to be sensitive to the deleterious effects of head injury if those with mTBI produced valid performance (Axelrod, Fichtenberg, Liethen, Czarnota,

& Stucky, 2001; Donders & Strong, 2015; Lange & Iverson, 2005).

Although this suggests that TMT validity cutoffs, especially the derivative indices, may be less sensitive to aspects of invalid performance that are responsible for the commonly observed reverse dose-response relationship, they also have the benefit of not being sensitive to genuine impairment. This finding is consistent with previous reports (Arnold et al., 2005; Erdodi et al., 2017; Erdodi, Hurtubise, et al., 2018; Erdodi et al., 2018; Erdodi, Tyson, et al., 2018; Jasinski, Berry, Shandera, & Clark, 2011). The variability in the signal detection profile across samples occasionally leads to accidental discoveries that reaffirm the utility of derivative EVIs (Axelrod, Meyers, & Davis, 2014; Dean, Victor, Boone, Philpott, & Hess, 2009; Glassmire, Wood, Ta, Kinney, & Nitch, 2019). Therefore, the cumulative evidence suggests that experimenting with derivative EVIs is a worthwhile exercise, even if they occasionally underperform.

The modality specificity effect (Erdodi, 2019; Rai and Erdodi, 2019) was observed in this study. The various cutoffs on the TMT were generally much more sensitive when the criterion PVT was the EI-5<sub>VM</sub>, which is composed of several visuomotor processing speed tests, compared to the WMT and the EI-5<sub>FCR</sub>, which are based on forced choice recognition paradigms. This finding underscores the importance of using a diverse combination of PVTs that vary in terms of cognitive domain and paradigm, to enable assessors to detect various manifestations of non-credible responding (Cottingham et al., 2014). In addition, the development of these EVIs will benefit clinicians because they provide non-memory-based EVIs to complement the commonly used PVTs that are typically based on forced choice recognition memory.

Overall, results suggest that the demographically corrected T-score cutoffs show promise in clinical use. These cutoffs had reasonably strong classification accuracy against the criterion PVTs and successfully eliminated potential age and education bias observed in the raw score cutoffs. TMT validity cutoffs were unrelated to injury severity, suggesting that they are robust to genuine cognitive impairment. This is particularly important given that the TMT is known to be sensitive to the deleterious effects of TBI (Sanchez-Cubillo et al., 2009; Woods et al., 2015).

Inevitably, the present study has a number of limitations, the most obvious of which is the relatively small and geographically restricted sample. The demographically corrected T-score cutoffs should be examined in other clinical populations and in other geographic regions to ensure the generalizability of findings. These cutoffs should also be examined with larger samples and using different criterion measures as well as research designs, including experimental malingering paradigms.

This study makes a small, but important contribution to the knowledge base of clinical neuropsychology. The use of PVTs

that are biased by demographic variables can lead to increased misclassification and contribute to erroneous diagnoses, especially in vulnerable populations (e.g., the elderly and examinees with lower levels of educational achievement). Ensuring that tests are not biased against demographic groups is a fundamental expectation in behavioral sciences. Violating this assumption can have serious social and ethical implications. In addition, in an era of increased financial and time pressures being placed on clinicians, establishing EVIs with strong classification accuracy may help to shorten evaluation time while simultaneously improving the assessment of performance validity (Boone, 2013; Erdodi, Pelletier, & Roth, 2018d; Larrabee, 2008, 2012).

### Compliance with Ethical Standards

Relevant ethical guidelines were followed throughout the project. All data collection, storage and processing was done with the approval of relevant institutional authorities regulating research involving human participants, in compliance with the 1964 Helsinki Declaration and its subsequent amendments or comparable ethical standards.

**Funding Information** This project received financial support from the University of Windsor, Faculty of Arts, Humanities, and Social Sciences Collaborative Research Grant.

**Conflict of Interest** This project received financial support from the University of Windsor, Faculty of Arts, Humanities, and Social Sciences Collaborative Research Grant. Relevant ethical guidelines were followed throughout the project. Drs. Abeare, Holcomb and Erdodi perform forensic consultations and medicolegal evaluations, for which they receive financial compensation.

### References

- Abeare, C., Messa, I., Whitfield, C., Zuccato, B., Casey, J., Rykalski, N., & Erdodi, L. (2018). Performance validity in collegiate football athletes at baseline neurocognitive testing. *The Journal of Head Trauma Rehabilitation*, 1.
- 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. <https://doi.org/10.1001/jamaneurol.2018.0031>.
- American Academy of Clinical Neuropsychology. (2007). American Academy of clinical neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*, 21(2), 209–231. <https://doi.org/10.1080/13825580601025932>.
- 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., Charles, J., Ali, S., Enache, A., Dhuga, J., & Erdodi, L. A. (2019). Re-examining 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. <https://doi.org/10.1080/13803395.2018.1483488>.
- 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. K., & Sugarman, M. A. (2017). Performance validity and processing speed in a VA Polytrauma sample. *The Clinical Neuropsychologist*, 31(5), 857–866.
- Ashendorf, L., Jefferson, A. L., O'Connor, M. K., Chaisson, C., Green, R. C., & Stern, R. A. (2008). Trail making test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology*, 23, 129–137.
- Axelrod, B. N., Fichtenberg, N. L., Liethen, P. C., Czarnota, M. A., & Stucky, K. (2001). Performance characteristics of postacute traumatic brain injury patients on the WAIS-III and WMS-III. *The Clinical Neuropsychologist*, 15(4), 516–520.
- Axelrod, B. N., Fichtenberg, N. L., Millis, S. R., & Wertheimer, J. C. (2006). Detecting incomplete effort with digit span from the Wechsler adult intelligence scale – Third edition. *The Clinical Neuropsychologist*, 20(3), 513–523.
- 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>.
- Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18, 632–642.
- Bigler, E. D. (2015). Neuroimaging as a biomarker in symptom validity and performance validity testing. *Brain Imaging and Behavior*, 9, 421–444. <https://doi.org/10.1007/s11682-015-9409-1>.
- 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, NY: 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>.
- 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.
- Busse, M., & Whiteside, D. (2012). Detecting suboptimal cognitive effort: Classification accuracy of the Conner's continuous performance test-II, brief test of attention, and trail making test. *The Clinical Neuropsychologist*, 26(4), 675–687.
- Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2013). Trail making test: Regression-based norms for the Portuguese population. *Archives of Clinical Neuropsychology*, 28(2), 189–198.
- 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.
- 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.
- Conners, K. C. (2004). *Conner's continuous performance test (CPT II). Version 5 for windows. Technical guide and software manual*. North Tonawanda, NY: Multi-Health Systems.

- Corrigan, J. D., & Hinkley, N. S. (1987). Relationships between parts a and B of the trail making test. *Journal of Clinical Psychology, 43*, 402–409.
- 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>.
- 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>.
- Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist, 23*, 133–152.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. (2000). *The California verbal learning test-second edition*. San Antonio TX: The Psychological Corporation.
- Donders, J., & Strong, C. A. H. (2015). Clinical utility of the Wechsler adult intelligence scale – Fourth edition after traumatic brain injury. *Assessment, 22*(1), 17–22.
- 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.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. San Antonio, TX: 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., 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.
- 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., Jongsma, K. A., & Issa, M. (2017). The 15-item version of the Boston naming test as an index of English proficiency. *The Clinical Neuropsychologist, 31*(1), 168–178. <https://doi.org/10.1080/13854046.2016.1224392>.
- 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. (2019). Information processing speed tests as PVTs. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment. A neuropsychological perspective*. New York, NY: Guilford.
- 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., Kirsch, N. L., Lajiness-O'Neill, R., & Medoff, B. (2014). 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., Sagar, S., Seke, K., Zuccato, B. G., Schwartz, E. S., & Roth, R. M. (2018). 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., 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 & Neuroscience, 10*(3), 325–344. <https://doi.org/10.1037/pne0000103>.
- 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., Tyson, B. T., Abeare, C. A., Zuccato, B. G., Rai, J. K., Seke, K. R., et al. (2018). 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.
- Gladis, 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.
- Glassmire, D. M., Wood, M. E., Ta, M. T., Kinney, D. I., & Nitch, S. R. (2019). Examining false-positive rates of Wechsler adult intelligence scale (WAIS-IV) processing speed based embedded validity indicators among individuals with schizophrenia spectrum disorders. *Psychological Assessment, 31*(1), 120–125. <https://doi.org/10.1037/pas0000650>.
- Golden, C. J., & van den Broek, A. (1998). Potential impact of age-and education-corrected scores on HRNB score patterns in participants with focal brain injury. *Archives of Clinical Neuropsychology, 13*(8), 683–694.
- Green, P. (2003). *Green's Word Memory Test*. Edmonton, Canada: Green's Publishing
- 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.
- Green, P., Iverson, G., & Allen, L. (1999). Detecting malingering in head injury litigation with the word memory test. *Brain Injury, 13*, 813–819.

- 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.
- Greve, K. W., Curtis, K. L., Bianchini, K. J., & Ord, J. S. (2009). Are the original and second edition of the California verbal learning test equally accurate in detecting malingering? *Assessment*, 16(3), 237–248.
- Grimes, D. A., & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *The Lancet*, 365(9469), 1500–1505.
- Halstead, W. (1947). *Brain and intelligence. A quantitative study of the frontal lobes*. Chicago: University of Chicago Press.
- Hamdan, A. C., & Hamdan, E. M. (2009). Effects of age and education on the trail making test in a healthy Brazilian sample. *Psychology & Neuroscience*, 2(2), 199–203.
- Hashimoto, R., Meguro, K., Lee, E., Kasai, M., Ishii, H., & Yamaguchi, S. (2006). Effect of age and education on the trail making test and determination of normative data for Japanese elderly people: The Tajiri project. *Psychiatry and Clinical Neurosciences*, 60, 422–428.
- Hayward, L., Hall, W., Hunt, M., & Zubrick, S. R. (1987). Can localized brain impairment be simulated on neuropsychological test profiles? *Australian and New Zealand Journal of Psychiatry*, 21, 87–93.
- 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, FL: 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, FL: Psychological Assessment Resources.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Participants, C. (2009). American Academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093–1129. <https://doi.org/10.1080/13854040903155063>.
- Heinly, M. T., Greve, K. W., Bianchini, K., Love, J. M., & Brennan, A. (2005). WAIS digit-span-based indicators of malingered neurocognitive dysfunction: Classification accuracy in traumatic brain injury. *Assessment*, 12(4), 429–444.
- Hester, R. L., Kinsella, G. J., Ong, B., & McGregor, J. (2005). Demographic influences on baseline and derived scores from the trail making test in healthy and older Australian adults. *The Clinical Neuropsychologist*, 19, 45–54.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295–300.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.
- 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.
- Iverson, G. L., Lange, R. T., Green, P., & Frazen, M. D. (2002). Detecting exaggeration and malingering with the trail making test. *The Clinical Neuropsychologist*, 16(3), 398–406.
- Jasinski, L. J., Berry, D. T., Shandera, A. L., & Clark, J. A. (2011). Use of the Wechsler adult intelligence scale digit span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology*, 33(3), 300–314.
- 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>.
- 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.
- Lange, R. T., Iverson, Z. K., M. L., Ethel-King, P. E., & Franzen, M. D. (2005). Interpreting the trail making test following brain injury: Comparison of traditional time scores and derived indices. *Journal of Clinical and Experimental Neuropsychology*, 27, 897–906.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17(3), 410–425. <https://doi.org/10.1076/clin.17.3.410.18089>.
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22, 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., 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.
- 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, 426–440.
- MacNeill Horton, A., & Roberts, C. (2001). Derived trail making test indices in a sample of substance abusers: Demographic effects. *International Journal of Neuroscience*, 1(1–2), 123–132.
- Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist*, 24, 1204–1237.
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of north American professionals. *The Clinical Neuropsychologist*, 29(6), 741–746.
- McCrea, M. A. (2008). *Mild traumatic brain injury and postconcussion syndrome: The new evidence base for diagnosis and treatment*. New York: Oxford University Press.
- Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, 29(3), 208–318.
- Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. New York, NY: Oxford University Press.
- Newcombe, F. (1969). *Missile wounds of the brain*. London: Oxford University Press.
- O'Bryant, S. E., Hilsabeck, R. C., Fisher, J. M., & McCaffrey, R. J. (2003). Utility of the trail making test in the assessment of malingering in a sample of mild traumatic brain injury litigants. *The Clinical Neuropsychologist*, 17(1), 69–74.
- 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, TX: Author.
- Perianez, J. A., Rios-Lago, M., Rodriguez-Sanchez, J. M., Adrover-Roig, D., Sanchez-Cubillo, I., Crespo-Facorro, B., et al. (2007). Trail making test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Archives of Clinical Neuropsychology*, 22, 433–447.
- Powell, M. R., Locke, D. E. C., Smigielski, J. S., & McCrea, M. (2011). Estimating the diagnostic value of the trail making test for suboptimal effort in acquired brain injury rehabilitation patients. *The Clinical Neuropsychologist*, 25(1), 108–118.
- 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., An, K. Y., Charles, J., Ali, S., & Erdodi, L. A. (2019). Introducing a forced choice recognition trial to the Rey Complex Figure Test. *Psychology & Neuroscience*.
- Rasmussen, X. D., Zonderman, A. B., Kawas, C., & Resnick, S. M. (1998). Effects of age and dementia on the trail making test. *The Clinical Neuropsychologist*, 12(2), 169–178.
- 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., & Wolfson, D. (1993). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (2nd ed.). Tucson, AZ: Neuropsychology Press.
- Retain, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology*, 19, 393–394.
- Ruffolo, L. F., Guilmette, T. J., & Willis, W. J. (2000). Comparison of time and error rates on the trail making test among patients with head injuries, experimental malingerers, patients with suspected effort on testing, and normal controls. *The Clinical Neuropsychologist*, 14, 223–230.
- Salazar, X. F., Lu, P. H., Wen, J., & Boone, K. B. (2007). The use of effort tests in ethnic minorities and in non-English speaking and English as a second language populations. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 405–427). New York: Guilford.
- Sanchez-Cubillo, I., Perianez, J. A., Adrover-Roig, D., Rodriguez-Sanchez, J. M., Rios-Lago, M., Tirapu, J. E. E. A., & Barcelo, F. (2009). Construct validity of the trail making test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15(3), 438–450.
- 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>.
- 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>.
- Solomon, R. E., Boone, K. B., Miora, D., Skidmore, S., Cottingham, M., Victor, T., et al. (2010). Use of the WAIS-III picture completion subtest as an embedded measure of response bias. *The Clinical Neuropsychologist*, 24, 1243–1256.
- Spencer, R. J., Axelrod, B. N., Drag, L. L., Waldron-Perrine, B., Pangilinan, P. H., & Bieliauskas, L. A. (2013). WAIS-IV reliable digit span is no more accurate than age corrected scaled score as an indicator of invalid performance in a veteran sample undergoing evaluation for mTBI. *The Clinical Neuropsychologist*, 27(8), 1362–1372.
- Spreen, O., & Benton, A. L. (1965). Comparative studies of some psychological tests for cerebral damage. *Journal of Nervous and Mental Disease*, 140, 323–333.
- Sugarman, M. A., & Axelrod, B. N. (2015). Embedded measures of performance validity using verbal fluency tests in a clinical sample. *Applied Neuropsychology: Adult*, 22(2), 141–146.
- 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>.
- Tombaugh, T. (2004). Trail making test a and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2), 203–214.
- 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.
- 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.
- Webber, T. A., Critchfield, E. A., & Soble, J. R. (2018). Convergent, discriminant and concurrent validity of nonmemory-based performance validity tests. Epublished ahead of print. *Assessment*, 107319111880487. <https://doi.org/10.1177/1073191118804874>.
- Wechsler, D. (2008). *Technical and interpretive manual for the Wechsler adult intelligence scale* (Fourth ed.). San Antonio, TX: Pearson.
- Wechsler, D. (2009). *Wechsler memory scale — Fourth edition (WMS-IV)*. San Antonio, TX: Pearson.
- 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.
- Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test 4*. Lutz, FL: Psychological Assessment Resources, Inc..
- 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.
- Woods, D. L., Wyma, J. M., Herron, T. J., & Yund, E. W. (2015). The effects of aging, malingering, and traumatic brain injury on computerized trail-making test performance. *PLoS One*, 10(6), e0124345.
- Yuspeh, R. L., Drane, D. L., Huthwaite, J. S., & Klingler, L. K. (2000). Derived Trail making test indices: Normative observations. *Archives of Clinical Neuropsychology*, 15, 653–850.

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