Detecting Negative Response Bias Within the Trauma Symptom Inventory–2 (TSI-2): a Review of the Literature

Francesca Ales[1] · Laszlo Erdodi[2]

© The Author(s) 2021

Abstract

This systematic review was performed to summarize existing research on the symptom validity scales within the Trauma Symptom Inventory–Second Edition (TSI-2), a relatively new self-report measure designed to assess the psychological sequelae of trauma. The TSI-2 has built-in symptom validity scales to monitor response bias and alert the assessor of non-credible symptom profiles. The Atypical Response scale (ATR) was designed to identify symptom exaggeration or fabrication. Proposed cutoffs on the ATR vary from ≥ 7 to ≥ 15, depending on the assessment context. The limited evidence available suggests that ATR has the potential to serve as measure of symptom validity, although its classification accuracy is generally inferior compared to well-established scales. While the ATR seems sufficiently sensitive to symptom over-reporting, significant concerns about its specificity persist. Therefore, it is proposed that the TSI-2 should not be used in isolation to determine the validity of the symptom presentation. More research is needed for development of evidence-based guidelines about the interpretation of ATR scores.

Keywords Trauma Symptom Inventory-2 · TSI-2 · Posttraumatic stress disorder · PTSD · Malingering · Symptom validity assessment

Background

The Trauma Symptom Inventory–Second Edition (TSI-2; Briere, 2011), the revised version of the original TSI (Briere, 1995), is a broadband self-report inventory designed to evaluate symptoms of posttraumatic stress disorder (PTSD) and of other non-specific psychological sequelae of traumatic events (e.g., insecure attachment styles, reduced self-capacity, and dysfunctional behaviors). It includes a wide range of complex acute and chronic symptoms, ranging from dissociation to somatization. Further, the triggering events that generated the trauma can be of various types and magnitude, such as sexual and physical assault, domestic and intimate partner violence, physical confrontation, combat, torture, motor vehicle collisions, major medical procedures, traumatic loss, and early experiences of childhood neglect or abuse. As such, the instrument was designed to assess a wide spectrum of trauma-induced symptoms across the adult lifespan in various clinical settings and contexts (e.g., inpatient/ outpatient, clinical/forensic etc.). However, as completely dependent on self-report, PTSD claims are not easy to assess, especially considering malingering as an issue.

The most recent edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines malingering as the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" (American Psychiatric Association, 2013, p. 726). Malingering is often exhibited as a strong tendency to claim one's health in a worse way than it actually is (i.e., negative response bias), leading to an over-endorsement of items included in symptom inventories. Malingering is estimated to costs millions of dollars each year and poses significant challenges to the mental health system (Chafetz & Underhill, 2013). Therefore, monitoring non-credible presentation has been identified as an important task during all psychological evaluations (Bush et al., 2014; Heilbronner et al., 2009; Young, 2014). In particular, PTSD is one of the most frequently

✉ Francesca Ales
francesca.ales@unito.it

Laszlo Erdodi
lerdodi@uwindsor.ca

1 Department of Psychology, University of Turin, Via Verdi 10, 10123 Turin, TO, Italy

2 Department of Psychology, University of Windsor, Windsor, ON, Canada

Springer

feigned disorders, especially in civil litigation (Hall & Hall, 2006). Therefore, non-credible presentation should be considered as a contributing factor during the assessment of PTSD.

There are several reasons why the PTSD diagnosis seems to be so vulnerable to malingering. First, since the DSM criteria emphasize subjective experiences, during the psychological evaluation, clinicians usually rely on the client's report (Guriel & Fremouw, 2003). Thus, the subjective nature of PTSD makes it a relatively easy disorder to feign in the absence of objective methods to cross-validate symptoms (Elhai et al., 2001; Frueh & Kinder, 1994; Lees-Haley, 1986). Second, claims of PTSD are often made in the context of external incentives to appear impaired (e.g., disability compensation, personal injury litigation, return to work evaluations, determining fitness for military duty), which may precipitate symptom fabrication or exaggeration or attribute pre-existing conditions to a new, compensable trauma. Hall and Hall (2006) report that in the 1990s, 14% of work injury claims were based on PTSD or other stress-related diagnoses. Frueh et al. (2005) report an estimate of over 40% malingered PTSD cases in the military, and Dube and Sadoff (2015) argue that indeed it would be relatively easy for a veteran to fabricate PTSD symptoms. Demakis and Elhai (2011) estimate a base rate of malingering PTSD of about 50%, and Suhr (2015) reports that at least 20% of PTSD symptoms are non-credible.

Third, PTSD is often comorbid with other preexisting conditions or personality disorders, making it difficult to distinguish between feigned versus genuine PTSD symptoms and equally importantly, the *cause* of the credible ones (compensable injury or long-standing, unresolved trauma history unrelated to the event in question). The potential coexistence of legitimate preexisting PTSD and symptom exaggeration/re-attribution further complicates the differential diagnosis (Elhai et al., 2001; Hyer et al., 1987). All of these factors make assessing the credibility of PTSD symptoms quite a challenge. Therefore, having well-calibrated psychometric tools to distinguish between bona fide and non-credible PTSD is of paramount importance.

The TSI-2 was developed in light of the fact that the first version of the TSI (Briere, 1995) was unable to detect instances when PTSD symptoms were over-reported, grossly exaggerated, or fully fabricated (Palermo & Brand, 2019). The TSI-2 consists of 136 items to which an examinee is asked to rate the frequency of their symptoms, condition, or behavior within the six-month period prior to the assessment. The response scale contains 4 points, from 0 ("never") to 3 ("all the time"). The cutoffs recommended in the Manual (Briere, 2011) are ≥ 8 for the general population, and ≥ 15 for clinical and forensic populations. Scores above the threshold may reflect either symptoms fabrication/exaggeration or authentic reporting of a person in a severe psychopathological state. For more details on the implications

of the cutoffs, see the "Cut Scores and Classification Accuracy" section.

TSI-2 scales and subscales assess a wide range of symptoms, such as intrusive experiences (e.g., flashbacks, nightmares, upsetting memories), anxiety, autonomic hyperarousal, and defensive avoidance (cognitive and behavioral) of distress. These symptoms are synchronized with Criteria B, C, and E of the DSM-5 (American Psychiatric Association, 2013). Depression, anger, irritability, disruptions in cognitive functioning, unusual behaviors, general health preoccupations, dissociation, sexual disturbances, identity confusion, and tension reduction behavior (e.g., using drugs, sexual acting out in order to reduce internal pain) are assessed as well. The TSI-2 also includes eight critical items related to dysfunctional behaviors that underlie severe psychological disturbances and represent indicators of danger to self and/or others (e.g., "Feeling the urge to physically hurt myself," "Considering to seriously hurt others"[1]).

All the items are divided into 12 clinical scales, some of which are in turn divided into subscales. The 12 clinical scales are grouped into four factors: (1) self-disturbance (SELF), (2) posttraumatic stress (TRAUMA), (3) externalizing (EXT), and (4) somatization (SOMA) (see Table 1 for a more thorough representation of factors, scales, and subscales).

Finally, the TSI-2 provides two symptom validity scales: the response level (RL), which assesses the tendency to deny common problems or under-report symptoms that others readily acknowledge, and the atypical responses (ATR), which assesses the tendency to exaggerate trauma-related symptoms. The 8-item ATR scale contains symptoms rarely endorsed by individuals with genuine trauma history, and has been redesigned specifically for the TSI-2 to improve the identification of non-credible response sets. Specifically, the ATR scale was refashioned so to assess not only exaggeration of symptoms, but also an inaccurate representation of PTSD symptomatology. Its detection mechanism is based on the combination of indiscriminate endorsement of extreme levels of genuine symptoms (i.e., the method of threshold) as well as bizarre symptoms that are rarely endorsed by patients with bona fide PTSD (Briere, 2011). In other words, both over-endorsement of common symptoms and endorsement of rare symptoms could result in failing the ATR scale.

Consequently, high scores on the ATR scale may indicate high levels of genuine distress, but also random responding or non-credible presentation (i.e., malingered PTSD). Therefore, the clinical utility of the instrument may be compromised by the fact that the TSI-2 assesses not only extreme levels of PTSD symptoms, dysregulation, insecure

---

[1] These are not actual TSI-2 items; they were generated by the authors of this paper to provide examples while also protecting test security.

Table 1 TSI-2 factors, scales, and subscales

Factor	Scale	Subscale
Self-disturbance (SELF)	*Depression (D)*	
	Insecure Attachment (IA)	Relational Avoidance (IA-RA)
		Rejection Sensitivity (IA-RS)
	Impaired Self-Reference (ISR)	Reduced Self-Awareness (IRS-RSA)
		Other-Directedness (IRS-OD)
Posttraumatic stress (TRAUMA)	*Anxious Arousal (AA)*	Anxiety (AA-A)
		Hyperarousal (AA-H)
	Intrusive Experience (IE)	
	Defensive Avoidance (DS)	
	Dissociation (DIS)	
Externalitation (EXT)	*Anger (ANG)*	
	Sexual Disturbance (SXD)	Sexual Concerns (SXD-SC)
		Dysfunctional Sexual Behavior (SXD-DSB)
	Suicidality (SUI)	Ideation (SUI-I)
		Behavior (SUI-B)
	Tension Reduction Behavior (TRB)	
Somatization (SOMA)	*Somatic Preoccupation (SOM)*	Pain (SOM-P)
		General (SOM-G)

attachment, and somatization, but also the exaggeration or outright fabrication of such symptoms. The fact that non-credible presentation and genuine elevation in symptoms can both inflate the score on the ATR scale poses significant conceptual and psychometric challenges in the standardization and clinical interpretation of the TSI-2 scores.

Validation and Psychometric Properties

The TSI-2 has been validated and normed on a sample of 678 adults whose demographic characteristics resembled those of the US Census (2007) for sex (54% women), age ($M = 53.4$, $SD = 18.3$, range = 18–90), ethnicity (73% Caucasian, 11% African American, 9% Hispanic, 7% other ethnic groups), education level, and geographic region. Reliability and validity were tested by employing five different (non-overlapping) clinical samples: patients with borderline personality disorder, survivors of domestic violence, survivors of sexual abuse, war veterans, and women in correctional facilities. In addition, to test the ability of the instrument to detect negative response bias, a non-clinical community sample was instructed to simulate PTSD (Briere, 2011). TSI-2 demonstrated strong psychometric properties: high internal consistency ($\alpha = 0.76$–0.94), and excellent test–retest reliability after an approximately one week interval ($r = 0.76$–0.93; Briere, 2011).

The first version of the ATR scale was developed based on the author's judgment of eccentricity and bizarreness

(Briere, 1995), and thus, it did not include data on criterion validity. Nevertheless, there was evidence of convergent validity: modest correlations ($r = 0.50$ and 0.52, respectively) with the F scale of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 1989), and the NIM scale of the Personality Assessment Inventory (PAI; Morey, 1991). The second version of the ATR scale has been considerably revised in order to specifically address issues regarding the utility of the first version in detecting negative response bias in forensic evaluations settings. In fact, instead of being based on general bizarre or "extreme" symptomatology as was the case for the ATR scale in the first version of the TSI, the ATR scale of the TSI-2 includes items that appear to indicate posttraumatic stress, but are rarely endorsed by individuals who genuinely suffer from PTSD.

In addition to evaluating internal consistency and temporal stability during instrument development, subsequent research also examined the TSI-2's concurrent, discriminant, construct, and criterion validity. However, recent studies (Palermo & Brand, 2019) have shown that the ATR scale, when used to distinguish coached simulators from patients with dissociative disorders—and using a cutoff of ≥ 15—underperforms compared to the Infrequency-Psychopathology (F_p) scale of the MMPI-2 (Butcher et al., 1989). In fact, the overall diagnostic power (ODP; percent of the sample correctly classified as having or not having the condition of interest) was 60% for the TSI-2 ATR scale, and 83% for the MMPI F_p scale (Brand & Chasson, 2015). Similarly, the Trauma Index of the Structured Interview of Reported Symptoms-2 (SIRS-2; Rogers et al., 1992), performed better (83%) than the ATR

scale (Brand et al., 2014). These findings are consistent with the fact that dissociative symptoms often overlap with the tendency to overreport bizarre and eccentric symptoms (Merckelbach et al., 2017).

Therefore, to determine the credibility of a given presentation, Palermo and Brand (2019) suggested that the TSI-2 should be used in conjunction with established symptom validity tests (SVTs), such as the SIRS-2 Trauma Index or the MMPI-2 F_p scale, given its incremental utility. In fact, previous studies have shown that these measures are effective at differentiating between genuine complex dissociative disorders and feigned symptoms (Brand & Chasson, 2015; Brand et al., 2014, 2016). Otherwise, the inherent limitations of the ATR scale may produce an unacceptably high false negative or false positive rate. On the other hand, recent studies on feigned PTSD showed that when combined with other instruments, the TSI-2 demonstrated incremental utility in distinguishing feigners from honest responders (Efendov et al., 2008; Elhai et al., 2005). Gray et al. (2010) compared a group of coached PTSD simulators with a group of patients with credible PTSD on the PAI (Morey, 1991) validity scales and the TSI-2 ATR scale. Both measures successfully differentiated between the two groups. However, the ATR scale did not perform as well as the positive impression management (PIM) and the Negative Impression Management (NIM) scales (Cohen's d: 1.10, 2.15, 2.19 for ATR, PIM_{PAI}, and NIM_{PAI}, respectively). At the same time, the ATR and NIM_{PAI} had comparable ODP (75% and 80%, respectively), which is the more clinically relevant parameter.

In a more recent study, Weiss and Rosenfeld (2017) compared the classification accuracy of three different validity tests in addition to the TSI-2: the Dot Counting Test (DCT; Boone et al., 2002), the Miller Forensic Assessment of Symptoms (M-FAST; Miller, 2001), and the Test of Memory Malingering (TOMM; Tombaugh, 1996) in a sample of trauma-exposed African immigrants. Three out of four tests (i.e., M-FAST, TSI-2 ATR scale, and DCT) effectively differentiated between participants instructed to feign and honest responders. The M-FAST and the TSI-2 ATR scale showed a moderate classification accuracy (AUC = 0.77 and AUC = 0.74, respectively, $p < 0.001$), with the DCT performing slightly lower (AUC = 0.66, $p = 0.027$).

Cut Scores and Classification Accuracy

The ATR cutoff scores recommended in the TSI-2 Professional Manual (Briere, 2011) are ≥ 8 for general or student populations and ≥ 15 for clinical and forensic populations. Only 4.8% of the combined clinical validation sample ($n = 125$) scored ≥ 15 on the ATR, suggesting high (0.95) specificity. Although failure rate was higher (10.9%) within

the PTSD subsample ($n = 55$), suggesting that significant trauma history contaminates the ATR scale, it closely approximated the 0.90 specificity standard (Briere, 2011).

To our knowledge, the TSI-2's ability to detect feigned PTSD has been evaluated in two earlier studies. In the first study (Gray et al., 2010), the scores of 75 young adults instructed to simulate PTSD symptoms were compared to those of 49 individuals who actually exhibited PTSD symptomatology. Predictably, simulators over-endorsed PTSD items included in the Anxious Arousal ($d = 0.48$), Intrusive Experiences ($d = 1.29$), and Defensive Avoidance ($d = 0.73$) scales. The Dissociation Scale was not examined because it was still under development at the time. The new version of the ATR scale (i.e., revised for TSI-2) outperformed the original version in detecting feigned PTSD. The authors identified ≥ 7 as the optimal cutoff score, which was able to classify 74% of feigners and 77% of individuals with genuine PTSD. With the same cutoff, Gray et al. (2010) reported an ODP of 0.75. It is worth noting that the original version of the ATR scale correctly classified only 59% of feigners and non-feigners in a study with a similar methodology (Elhai et al., 2005).

In the second study, Weiss and Rosenfeld (2017) assessed the effect of demographic and cultural variables classification accuracy on several validity tests. They compared performance on the ATR validity scale among three groups of African immigrants (i.e., honest responders with PTSD vs. honest responders without PTSD vs. participants asked to feign PTSD-related symptoms). The ATR scale produced a significant AUC value (0.74). However, at the cutoff score of ≥ 15, the ATR scale had low sensitivity (0.32) at 0.90 specificity. The authors could not identify a more effective cutoff. Additionally, a recent study (Filone & DeMatteo, 2017) evaluated a sample of 97 individuals who claimed to be traumatized by child abuse, war, and torture. A small subset (6.2%) of them failed the cutoff score of ≥ 15. However, the study did not report data on objective measures of symptom validity, thus it is not known whether this failure rate should be construed as the false positive rate or the proportion of correctly detected invalid protocols.

Finally, Palermo and Brand (2019) examined the TSI-2 profile of individuals with complex dissociative disorders (CDDs). Given the relationship between exposure to severe trauma and the presence of CDD (Brand & Stadnik, 2013; Brand et al., 2013; Foote et al., 2008; Rodewald et al., 2011; Saxe et al., 2002), the authors hypothesized that patients with CDD would have scores comparable or higher than patients with PTSD. Thus, they compared coached CDD feigners with *bona fide* CDD profiles on the TSI-2 clinical and validity scales and examined the utility of the ATR in distinguishing credible from non-credible CDD. Classification accuracy was reported at three levels of ATR: the optimal cutoff score (≥ 7) reported by Gray et al. (2010), the original cutoff recommended by Briere for general populations (≥ 8), and the cutoff Briere (2011)

recommended for clinical or forensic populations (≥ 15). As expected, the most liberal cutoff (≥ 7) obtained high sensitivity (0.92) but had unacceptably low specificity (0.49), with an ODP of 0.73. Making the cutoff slightly more conservative (≥ 8) resulted in small improvement in specificity (0.51) at a disproportional cost to sensitivity (0.86) and an ODP of 0.71. Raising the cutoff to (≥ 15) sacrificed much of the sensitivity (0.47) for a meaningful increase in specificity (0.77). However, there was a notable decline in ODP (0.60).

These results are consistent with the recommendations of the TSI-2 Manual, which cautions against using a cutoff of ≥ 8 in clinical or forensic settings. The unacceptably high false positive rate (49%) at this cutoff substantiates the authors' concerns. However, even at the much more conservative cutoff score of ≥ 15, the false positive rate was still high (33%), even as sensitivity dropped to 47%. Moreover, ODP declined as the cutoff increased, indicating that even if sensitivity and specificity were sacrificed, a high false positive rate persisted. Results suggests that the ATR is more efficient at separating credible from non-credible profiles among individuals with PTSD symptoms (i.e., 89.5%; Weiss & Rosenfeld, 2017), and trauma histories (i.e., 93.8%; Filone & DeMatteo, 2017) rather than CDD (Palermo & Brand, 2019). Of course, these findings must be replicated before definitive population-specific recommendations can be issued. However, as anticipated earlier, it is worth mentioning that the link between dissociative symptoms and overreporting of bizarre symptoms has been previously identified as a potential confound (Merckelbach et al., 2017). In a recent critical review of case studies on dissociative amnesia, Mangiulli et al. (2021) report that the evidence of autobiographical memory loss was weak, and insufficient to establish dissociative amnesia. In fact, most of the cases examined in the study were not able to take into account other possible factors, such as ordinary forgetfulness or malingering.

Strength and Weaknesses

An apparent advantage of the TSI-2 is that it covers a wide range of potential sequelae of trauma history. The TSI-2 allows clinicians to generate a comprehensive symptom profile for patients who survived significant traumatic experiences. In addition, validity scales assess the tendency to exaggerate symptoms associate with trauma history. Moreover, the presence of some critical items and scales such as Suicidal Propensity, Sexual Disorders, or Behavior Aimed at Reducing Tension can alert the assessor to relevant clinical features that require immediate attention, especially with managing risk of suicide.

From a practical standpoint, its brevity (20 min) and self-administered nature are definite strengths of the TSI-2. The instrument can be administered in the traditional (paper-and-pencil) format or on a computer via PARiConnect, an online assessment platform. The latter is particularly relevant in the age of a pandemic where a significant proportion of clinical assessment is performed remotely. Scoring is equally fast and can be done (a) by hand in about 20 min, (b) using scoring software (TSI-2-SP), or (c) online 24/7 via PARiConnect.

Although not necessarily a weakness, the TSI-2 is primarily a measure of psychopathology, which provides two measures of symptoms validity. In contrast, more robust self-report inventories contain a variety of well-researched validity scales, such as the MMPI-2 (Butcher et al., 2001), the MMPI-2-RF (Ben-Porath & Tellegen, 2008), the PAI (Morey, 1991, 2007), the Millon Clinical Multiaxial Personality Inventory-IV (MCMI-IV; Millon et al., 2015), that do assess response style. Clinicians should consider this limitation of the TSI-2's effectiveness in determining the credibility of a given response set (Sellbom & Suhr, 2019). In fact, although the TSI-2 is an evidence-based measure, its validity scales fall short of the standards established by the most widely used Evidence-Based Psychological Assessment self-reports (e.g., PAI, MMPI-2-RF; Sellbom & Suhr, 2019), which may explain the generally low classification accuracy of the ATR. Specifically, the overlap between non-credible profiles and profiles of patients with bona fide severe symptoms is both a conceptual and practical barrier to the improvement of its signal detection performance. Finally, since the TSI-2 is a relatively new measure, more research is needed on its clinical utility to establish empirically based guidelines for the interpretation of its validity scales and symptom profiles.

Future Directions

To date, the TSI-2 is available in English and Spanish. Although there are no published standards yet for the Spanish translation, results from a sample of undergraduate students in Puerto Rico suggest good internal consistency, as well as good content validity (Gutiérrez Wang et al., 2011). However, more research is clearly needed to validate the test in other countries with different populations.

Overall, the TSI-2 would also benefit from more research on its classification accuracy across well-defined clinical populations with a wide range of demographic characteristics and geographic areas (Lichtenstein et al., 2019). The existing knowledge base on the ATR scale is limited by the reliance on simulation studies (Abeare et al., 2020; Carvalho et al., 2021; Hurtubise et al., 2020; Rai et al., 2019) and using performance validity tests[2] (as opposed to self-report symptom validity tests) as criterion measures (Gegner et al.,

---

[2] Larrabee (2012) recommends of two different terms in assessing the validity of an evaluee examination: performance validity and symptom validity. The former "refers to the validity of the actual ability to perform a task, assessed either by standalone tests or by atypical performance on neuropsychological tests"; the latter, instead "refers to the validity of symptomatic complaints on self-report measures" (Ales et al., 2021).

2021; Sabelli et al., 2021). Modality specificity is an increasingly recognized methodological artifact in calibrating instruments (Erdodi, 2021; Giromini et al., 2020; Lace et al., 2020; Schroeder et al., 2019). Therefore, cross-validating the ATR using strategically selected and well-established criterion measures is recommended before making high-stake decisions about the credibility of a given profile. Further empirical investigations are needed to determine the optimal cutoff on the ATR scales. Alternative validity scales based on different detection methods (response consistency, rarely endorsed symptoms, logistic regression equations) should also be considered. Finally, in an increasingly diverse world, studies examining the cross-cultural validity of the TSI-2 would greatly expand the scope of the instrument (Ali et al., 2020; Erdodi et al., 2017).

References

Abeare, C. A., Hurtubise, J., Cutler, L., Sirianni, C., Brantuo, M., Makhzoun, N., & Erdodi, L. (2020). Introducing a forced choice recognition trial to the Hopkins Verbal Learning Test – Revised. *The Clinical Neuropsychologist*. Advance online publication. https://doi.org/10.1080/13854046.2020.1779348

Ales, F., Giromini, L., Warmelink, L., Polden, M., Wilcockson, T., Kelly, C., Winters, C., Zennaro, A., & Crawford, T. (2021). An eye tracking study on feigned schizophrenia. *Psychological Injury and Law*. https://doi.org/10.1007/s12207-021-09421-1

Ali, S., Elliott, L., Biss, R., Abumeeiz, M., Brantuo, M., Kuzmenka, P., Odenigbo, P., & Erdodi, L. (2020). The BNT-15 provides an accurate measure of English proficiency in cognitively intact bilinguals – A study in cross-cultural assessment. *Applied Neuropsychology: Adult*. Advance online publication.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

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

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

Brand, B. L., & Chasson, G. S. (2015). Distinguishing simulated from genuine dissociative identity disorder on the MMPI-2. *Psychological Trauma: Theory, Research, Practice, and Policy, 7*, 93–101. https://doi.org/10.1037/a0035181

Brand, B. L., Chasson, G. S., Palermo, C. A., Donato, F. M., Rhodes, K. P., & Voorhees, E. F. (2016). Truth is in the details: A comparison of MMPI-2 item endorsements in dissociative identity disorder vs. simulators. *Journal of the American Academy of Psychiatry and the Law, 44*, 63–72.

Brand, B. L., McNary, S. W., Myrick, A. C., Classen, C. C., Lanius, R., Loewenstein, R. J., Pain, C., & Putnam, F. W. (2013). A longitudinal naturalistic study of patients with dissociative disorders treated by community clinicians. *Psychological Trauma: Theory, Research, Practice, and Policy, 5*, 301–308. https://doi.org/10.1037/a0027654

Brand, B. L., & Stadnik, R. (2013). What contributes to predicting change in the treatment of dissociation: Initial levels of dissociation, PTSD, or overall distress? *Journal of Trauma & Dissociation, 14*, 328–341. https://doi.org/10.1080/15299732.2012.736929

Brand, B. L., Tursich, M., Tzall, D., & Loewenstein, R. J. (2014). Utility of the SIRS-2 in distinguishing genuine from simulated dissociative identity disorder. *Psychological Trauma: Theory, Research, Practice, and Policy, 6*, 308–317. https://doi.org/10.1037/a0036064

Briere, J. (1995). *Trauma Symptom Inventory (TSI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Briere, J. (2011). *Trauma Symptom Inventory-2nd edition (TSI-2) professional manual*. Odessa, FL: Psychological Assessment Resources. https://doi.org/10.1002/9780470479216.corpsy1010

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*, 197–205. https://doi.org/10.1007/s12207-014-9198-7

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. University of Minnesota Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, G. (2001). *Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring* (2nd ed.). Minneapolis, MN: University of Minnesota Press.

Carvalho, L., Reis, A., Colombarolli, M. S., Pasian, S. R., Miguel, F. K., Erdodi, L. A., Viglione, D. J., & Giromini, L. (2021). Discriminating feigned from credible PTSD symptoms: A validation of a Brazilian version of the Inventory of Problems-29 (IOP-29). *Psychological Injury and Law, 14*, 58–70. https://doi.org/10.1007/s12207-021-09403-3

Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology, 28*, 633–639. https://doi.org/10.1093/arclin/act038

Demakis, G. J., & Elhai, J. D. (2011). Neuropsychological and psychological aspects of malingered posttraumatic stress disorder. *Psychological Injury and Law, 4*, 24–31. https://doi.org/10.1007/s12207-011-9099-y

Dube, A. R., & Sadoff, R. L. (2015). Veterans affairs entitlement for service-connected disability caused by posttraumatic stress disorder. *The Journal of the American Academy of Psychiatry and the Law, 43*(2), 239–240.

Efendov, A. A., Sellbom, M., & Bagby, R. M. (2008). The utility and comparative incremental validity of the MMPI-2 and Trauma Symptom Inventory validity scales in the detection of feigned PTSD. *Psychological Assessment, 20*, 317–326. https://doi.org/10.1037/a0013870

Elhai, J. D., Gold, P. B., Sellers, A. H., & Dorfman, W. I. (2001). The detection of malingered posttraumatic stress disorder with MMMPI-2 Fake Bad indices. *Assessment, 8*(2), 221–236. https://doi.org/10.1177/107319110100800210

Elhai, J. D., Gray, M. J., Naifeh, J. A., Butcher, J. J., Davis, J. D., Falsetti, S. A., & Best, C. L. (2005). Utility of the Trauma Symptom Inventory's Atypical Response Scale in detecting malingered post-traumatic stress disorder. *Assessment, 12*(2), 210–219. https://doi.org/10.1177/1073191105275456

Erdodi, L. A. (2021). *Five shades of gray: Conceptual and methodological issues around multivariate models of performance validity*. Advance online publication.

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.

Filone, S., & DeMatteo, D. (2017). Assessing "credible fear": A psychometric examination of the Trauma Symptom Inventory-2 in the context of immigration court evaluations. *Psychological Assessment, 29*, 701–709. https://doi.org/10.1037/pas0000477

Foote, B., Smolin, Y., Neft, D. I., & Lipschitz, D. (2008). Dissociative disorders and suicidality in psychiatric outpatients. *Journal of Nervous and Mental Disease, 196*, 29–36. https://doi.org/10.1097/NMD.0b013e31815fa4e7

Frueh, B. C., Elhai, J. D., Grubaugh, A. L., Monnier, J., Kashdan, T. B., Sauvageot, J. A., Hamner, M. B., Burkett, B. G., & Arana, G. W. (2005). Documented combat exposure of US veterans seeking treatment for combat-related post-traumatic stress disorder. *British Journal of Psychiatry, 186*(6), 467–472.

Frueh, B. C., & Kinder, B. N. (1994). The susceptibility of the Rorschach Inkblot Test to malingering of combat-related PTSD. *Journal of Personality Assessment, 62*(2), 280–298. https://doi.org/10.1207/s15327752jpa6202_9

Gegner, J., Erdodi, L. A., Giromini, L., Viglione, D. J., Bosi, J., & Brusadelli, E. (2021). An Australian study on feigned mTBI using the Inventory of Problems–29 (IOP-29), its Memory Module (IOP-M), and the Rey Fifteen Item Test (FIT). *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2020.1864375

Giromini, L., Viglione, D. J., Zennaro, A., Maffei, A., & Erdodi, L. A. (2020). SVT meets PVT: Development and initial validation of the Inventory of Problems – Memory (IOP-M). *Psychological Injury and Law, 13*, 261–274. https://doi.org/10.1007/s12207-020-09385-8

Gray, M. J., Elhai, J. D., & Briere, J. (2010). Evaluation of the Atypical Response scale of the Trauma Symptom Inventory-2 in detecting simulated posttraumatic stress disorder. *Journal of Anxiety Disorders, 24*, 447–451. https://doi.org/10.1016/j.janxdis.2010.02.011

Guriel, J., & Fremouw, W. (2003). Assessing malingered posttraumatic stress disorder: A critical review. *Clinical Psychology Review, 23*(7), 881–904. https://doi.org/10.1016/j.cpr.2003.07.001

Gutiérrez Wang, L., Cosden, M., & Bernal, G. (2011). Adaptation and validation of the Spanish-language Trauma Symptom Inventory in Puerto Rico. *Journal of Consulting and Clinical Psychology, 79*, 118–122. https://doi.org/10.1037/a0021327

Hall, R. C. W., & Hall, R. C. W. (2006). Malingering of PTSD: Forensic and diagnostic considerations, characteristics of malingerers and clinical presentations. *General Hospital Psychiatry, 28*, 525–535. https://doi.org/10.1016/j.genhosppsych.2006.08.011

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

Hurtubise, J., Baher, T., Messa, I., Cutler, L., Shahein, A., Hastings, M., Carignan-Querqui, M., & Erdodi, L. (2020). Verbal fluency and digit span variables as performance validity indicators in experimentally induced malingering and real world patients with TBI. *Applied Neuropsychology: Child, 9*(4), 337–354. https://doi.org/10.1080/21622965.2020.1719409

Hyer, L., Fallon, J. H., Harrison, W. R., & Boudewyns, P. A. (1987). MMPI overreporting by Vietnam combat veterans. *Journal of Clinical Psychology, 43*(1), 79–83. https://doi.org/10.1002/1097-4679(198701)43:1%3C79::AID-JCLP2270430110%3E3.0.CO;2-P

Lace, J. W., Grant, A. F., Kosky, K. M., Teague, C. L., Lowell, K. T., & Gfeller, J. D. (2020). Identifying novel embedded performance validity test formulas within the Repeatable Battery for the Assessment of Neuropsychological Status: A simulation study. *Psychological Injury and Law, 13*, 303–315. https://doi.org/10.1007/s12207-020-09382-x

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of International Neuropsychological Society, 18*(4), 625–630. https://doi.org/10.1017/S13556177120002

Lees-Haley, P. R. (1986). Pseudo post-traumatic stress disorder. *Trial Diplomacy Journal, 9*(4), 17–20.

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. *Psychological Injury and Law, 12*(2), 127–145. https://doi.org/10.1007/s12207-019-0935

Mangiulli, I., Otgaar, H., Jelicic, M., & Merckelbach, H. (2021). A critical review of case studies on dissociative amnesia. *Clinical Psychological Science*. https://doi.org/10.1177/21677026211018194

Merckelbach, H., Boskovic, I., Pesy, D., Dalsklev, M., & Lynn, S. J. (2017). Symptom overreporting and dissociative experiences: A qualitative review. *Consciousness and Cognition, 49*, 132–144. https://doi.org/10.1016/j.concog.2017.01.007

Miller, H. (2001). *Manual for the Miller Forensic Assessment of Symptoms Test (M-FAST)*. Odessa, FL: Psychological Assessment Resources.

Millon, T., Millon, C., & Grossman, S. (2015). *Millon Clinical Multiaxial Inventory – IV manual* (4th ed.). NCS Pearson.

Morey, L. (1991). *The Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (2007). *Personality assessment inventory (PAI): Professional manual* (2nd ed.). Psychological Assessment Resources.

Palermo, C. A., & Brand, B. L. (2019). Can the Trauma Symptom Inventory-2 distinguish coached simulators from dissociative disorder patients? *Psychological Trauma: Theory, Research, Practice, and Policy, 11*(5), 477–485. https://doi.org/10.1037/tra0000382

Rai, J. K., 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, 12*(4), 451–472. https://doi.org/10.1037/pne0000175

Rodewald, F., Wilhelm-Göling, C., Emrich, H. M., Reddemann, L., & Gast, U. (2011). Axis-I comorbidity in female patients with dissociative identity disorder and dissociative identity disorder not otherwise specified. *Journal of Nervous and Mental Disease, 199*, 122–131. https://doi.org/10.1097/NMD.0b013e318208314e

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured interview of reported symptoms: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Sabelli, A. G., Messa, I., Giromini, L., Lichtenstein, J. D., May, N., & Erdodi, L. A. (2021). Symptom versus performance in patients with mild TBI: Independent sources of non-credible responding. *Psychological Injury and Law, 14*(1), 17–36. https://doi.org/10.1007/s12207-021-09400-6

Saxe, G. N., Chawla, N., & van der Kolk, B. (2002). Self-destructive behavior in patients with dissociative disorders. *Suicide &*

Life-Threatening Behavior, 32, 313–320. https://doi.org/10.1521/suli.32.3.313.22174

Schroeder, R. W., Martin, P. K., Heindrichs, R. J., & Baade, L. E. (2019). Research methods in performance validity testing studies: Criterion grouping approach impacts study outcomes. *The Clinical Neuropsychologist, 33*(3), 466–477. https://doi.org/10.1080/13854046.2018.1484517

Sellbom, M., & Suhr, J. A. (Eds.). (2019). *The Cambridge Handbook of Clinical Assessment and Diagnosis*. Cambridge University Press.

Suhr, J. A. (2015). *Psychological assessment: A problem-solving approach*. Guilford Press.

Tombaugh, T. N. (1996). *TOMM: Test of Memory Malingering*. North Tonawanda, NY: Multi-Health Systems.

U.S. Census Bureau. (2007). *Current population survey, March 2007 [Data file]*. Washington, DC: U.S. Department of Commerce.

Weiss, R. A., & Rosenfeld, B. (2017). Identifying feigning in trauma-exposed African immigrants. *Psychological Assessment, 29*, 881–889. https://doi.org/10.1037/pas0000381

Young, G. (2014). *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court*. Springer.