Psychological Injury and Law
https://doi.org/10.1007/s12207-020-09373-y

Psychometric Performance of the Miller Forensic Assessment of Symptoms Test (M-FAST) in Veteran PTSD Assessment

Erika J. Wolf[1,2] · Stephanie Ellickson-Larew[1] · Rachel E. Guetta[1,3] · Shaline Escarfulleri[1] · Karen Ryabchenko[1,2] · Mark W. Miller[1,2]

Received: 30 October 2019 / Accepted: 1 March 2020

Abstract

This study examined the psychometric properties of a widely used measure of symptom exaggeration, the Miller Forensic Assessment of Symptoms Test (M-FAST, Miller, 2001), in a sample of 209 (83.7% male) trauma-exposed veterans (57.9% probable current posttraumatic stress disorder; PTSD). M-FAST total scores evidenced acceptable internal consistency, but several subscales showed poor internal consistency. Factor analytic and item-response theory analyses identified seven poorly performing items. Comparisons with other measures of psychopathology and response validity (including subscales from the Minnesota Multiphasic Personality Inventory-2 Restructured Form) revealed that M-FAST scores were highly correlated with indices of psychopathology while less strongly associated with measures of symptom over-reporting. Empirically and clinically-derived (using a follow-up testing-the-limits procedure) revised M-FAST scores failed to improve the measure's psychometric performance. Results raise concerns about the validity of the M-FAST for identifying malingering in veterans with PTSD and carry implications for access to care and forensic evaluations in this population.

Keywords Malingering · Symptom exaggeration · M-FAST · PTSD · Over-reporting · Psychometric

One of the most commonly used stand-alone measures of symptom exaggeration in psychiatric and disability-claim-related populations is the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001). The measure was developed initially in forensic inpatient psychiatry units to quickly screen for exaggerated symptom reporting consistent with malingering using an interview format in which a clinician reads a series of statements concerning unusual psychological symptoms and suggestibility and provides categorical response options for the respondent (e.g., true/false, always/sometimes/never), with each item scored dichotomously (present/absent). It was designed to align with the Structured Interview for Reported Symptoms (SIRS, Rogers, Bagby, & Dickens, 1992), a lengthier, structured clinician-administered interview that was considered to be the "gold standard" diagnostic interview for malingering (the updated version of which, the SIRS-2, is now considered the gold standard; Rogers, Sewell, & Gillard, 2010). The M-FAST includes seven subscales: Reported vs. Observed (symptoms that the individual endorses but that are not observed by the interviewer), Extreme Symptomatology (highly atypical symptoms), Rare Combinations (symptoms that although common individually, are unlikely to co-occur), Unusual Hallucinations (atypical descriptions of hallucinations), Unusual Symptom Course (unusual start/stop of symptom), Negative Image (extremely negative self-image), and Suggestibility (responding affirmatively to a leading statement by the interviewer concerning an odd psychological symptom). Miller (2001) aimed to develop items that were rarely endorsed, even in samples with psychosis, due to their extreme and unusual nature. The M-FAST has frequently been recommended for use for the assessment of malingering in posttraumatic stress disorder (PTSD) samples

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12207-020-09373-y) contains supplementary material, which is available to authorized users.

✉ Erika J. Wolf
erika.wolf@va.gov

[1] National Center for PTSD at VA Boston Healthcare System, 150 S. Huntington Ave. (116B-2), Boston, MA, USA

[2] Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA

[3] Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

Published online: 15 April 2020

(e.g., Ali, Jabeen, & Alam, 2015; Knoll & Resnick, 2006) because PTSD symptoms can potentially be feigned for secondary gain purposes in criminal or civil proceedings, as discussed in greater detail below (Arbisi, Ben-Porath, & McNulty, 2006; Frueh, Smith, & Barker, 1996; Rogers et al., 1992).

In developing the M-FAST, Miller (2001) examined a large pool of potential items and selected a final set of 25 based on their correlations with group membership (items with $r \geq .30$ were retained) in a simulation study in which undergraduates were instructed to either answer honestly to the items or to feign psychiatric symptoms. Multiple validation samples yielded evidence of strong total score and subscale internal consistency and test-retest reliability over 1–3 weeks (Miller, 2001). However, evaluation of the psychometric properties and scoring of the M-FAST in independent samples has tended to find more variable evidence for internal reliability and criterion validity. Guy and Miller (2004) reported acceptable coefficient alpha for overall M-FAST total scores (α = 0.90) in a sample of incarcerated males; however, the coefficient alpha for subscale scores were notably weaker, with some unacceptably low (α = 0.29–0.53). Other studies have evaluated samples in which a group of individuals was directed to feign symptoms of a particular psychiatric diagnosis on the M-FAST, and these scores were compared to those who were instructed to respond honestly to the interview, including several psychiatric patient groups. For example, Guy, Kwartner, and Miller (2006) compared the psychometric properties of the M-FAST for simulators and a variety of patient groups and found that internal consistency for the total scale was notably weaker in a PTSD sample (α = 0.47) relative to undergraduates who were instructed to feign PTSD (α = 0.83; Guy et al., 2006). In that study, coefficient alpha for subscale scores were weak in both the PTSD and PTSD simulator groups (αs = 0.27–0.76), with some values unreported in the PTSD group due to negative associations among the items.

Exploration of the structure of the measure using principal components analysis suggested that the items could be reduced to a single weighted component (Miller, 2001). Similarly, a follow-up study of two forensic samples (one of which overlapped that used in the exploratory study) found that a one-factor model provided the best fit to the data (Vitacco et al., 2008). However, the analysis also revealed that one item (no. 5) had an unacceptable loading of − 0.03 and the standardized root mean square residual (SRMR) fit index was out of bounds, suggesting model misspecification. Therefore, it is not clear that a one-factor structure best represents the data, suggesting the need for exploratory factor analyses to test the factor structure of the measure. With respect to criterion and construct validity, studies of criminal defendants and incarcerated males have shown that M-FAST total scores were strongly correlated with the SIRS total score, with categorical SIRS-based malingering classification (Guy & Miller, 2004; Miller, 2001, 2004), and with over-reporting scales on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Graham, Tellegen, & Kaemmer, 1989).

A cut score of 6+ on the M-FAST is commonly recommended as optimal for diagnostic accuracy, based on a variety of clinical (including forensic) and non-clinical samples (Guy & Miller, 2004; Miller, 2001). However, some studies of psychiatric symptoms have found this threshold to yield unacceptable classification profiles and have instead suggested higher thresholds, ranging from 8+ (Veazey, Wagner, Hays, & Miller, 2005) to 16+ (Glassmire, Tarescavage, & Gottfried, 2016). For example, Guy et al. (2006) evaluated M-FAST diagnostic efficiency among undergraduates assigned to feign psychiatric symptoms (including PTSD) and patients with true psychopathology (malingering in this group was ruled out via SIRS administration). They found that at the recommended threshold of 6+ on the M-FAST total score, sensitivity was weak (0.63), while specificity (0.85) and the total correct classification (0.72) were acceptable for distinguishing the PTSD versus simulator group. The classification accuracy was poorer in the PTSD group than in the other patient groups that were evaluated (depression, schizophrenia, bipolar). Even when using a higher M-FAST threshold (8+), one study of Iraq and Afghanistan war era veterans with PTSD found a surprisingly high prevalence of malingering (75%; Constans et al., 2014), raising doubt about the accuracy of this determination. In addition to concerns about false positives in PTSD samples, the M-FAST may also be susceptible to false negatives. This was evident in studies that reported that coaching undergraduates to feign PTSD symptoms in a more sophisticated manner resulted in a missed symptom over-reporting classification nearly half of the time (Guriel-Tennant & Fremouw, 2006; Guriel et al., 2004).

Additional evidence for psychological factors affecting the validity of M-FAST scores in PTSD samples comes from studies that have examined the correlates of the measure. A study of veterans of the wars in Iraq and Afghanistan examined how groups with and without PTSD and with and without symptom exaggeration (defined as 8+ on the M-FAST) responded to a modified Stroop task featuring trauma-related words, based on prior evidence that individuals with PTSD tend to respond slower to trauma-relevant words (Constans et al., 2014). Results suggested that those with PTSD who were identified as over-reporting on the M-FAST evidenced greater response latencies to trauma-relevant words, relative to those with PTSD who were not identified as over-reporting and relative to those without PTSD. Given that response latency is automatic and not prone to manipulation (Constans, McCloskey, Vasterling, Brailey, & Mathews, 2004), particularly to category-specific stimuli (Buckley, Galovski, Blanchard, & Hickling, 2003), this result was interpreted as evidence that the group identified as exaggerating simply had

greater psychological distress (Constans et al., 2014). Other features of psychopathology may also unduly influence scores on the measure: a study of military veterans with PTSD found that the association between PTSD and scores on the M-FAST was moderated by alexithymia, raising the possibility that lack of emotional understanding and language may contribute to high scores on the measure (Brady, Bujarski, Feldner, & Pyne, 2017). Collectively, these studies raise questions about the validity of the M-FAST for PTSD samples. They suggest the susceptibility of the measure to true psychopathology and to misinterpretation. Moreover, most studies concerning the use of the M-FAST to identify feigned PTSD have been conducted using a simulation design in which one group (often college students) was instructed to feign symptoms, and it is unclear to what extent results from these studies generalize to PTSD samples. Ensuring the appropriateness of the measure to the PTSD population is of critical importance given the potential implications of malingering determinations for treatment and compensation decisions.

The primary aim of this study was to examine the psychometric properties of the M-FAST in a veteran sample with a high prevalence of PTSD. To do so, we examined the reliability and factor structure of the measure when administered per manual scoring rules. We used the results of these analyses to identify the best items on the measure and eliminate poorly performing ones to test if doing so improved the psychometric properties of the measure compared to the original scale. We also examined the construct validity of the M-FAST. Our evaluation of construct validity focused on the use of the MMPI-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008; Tellegen & Ben-Porath, 2008), which has previously been used to identify over-reporting of psychiatric symptoms among PTSD samples and to identify PTSD-related comorbidity (Goodwin, Sellbom, & Arbisi, 2013; Marion, Sellbom, & Bagby, 2011). For example, the Fp-r scale (Infrequent Psychopathology Responses), which captures symptoms that are rarely endorsed even among psychiatric populations, has been shown to identify individuals instructed to feign symptoms of PTSD (Goodwin et al., 2013; Marion et al., 2011) and is only moderately elevated among individuals with the disorder (Arbisi, Polusny, Erbes, Thuras, & Reddy, 2011). The breadth of the MMPI-2-RF Higher-Order and Restructured Clinical scales, spanning both internalizing and externalizing personality pathology, provided for a comprehensive investigation of the association between psychopathology and malingering-related scores in this sample.

Given concerns that the M-FAST may be sensitive to true psychopathology and lack of item clarity (e.g., Constans et al., 2014; Weiss & Rosenfeld, 2017), a second aim was to conduct a testing-the-limits follow-up interview to determine if doing so would clarify responses, yield information concerning item-level validity, and inform a revised total score to better index malingering. Testing-the-limits (Lezak, Howieson, Loring, & Fischer, 2004) is a common psychological assessment technique in which a clinician further queries an examinee's responses to test items to encourage the individual to elaborate on the initial response and/or provide responses to items the examinee might have otherwise not been able to complete (e.g., because the examinee timed or tested out, which can occur in intelligence and aptitude testing). It can be used when there is concern that test scores may not reflect true ability or the underlying construct (Lezak et al., 2004). We hypothesized that the scores on the original M-FAST would show strong associations with psychiatric distress and symptoms, as defined by the MMPI-2-RF Higher-Order scales indexing internalizing, externalizing, and thought disorder symptoms, which would diminish the validity of the measure for identifying malingering in particular, and result in relatively weaker associations with MMPI-2-RF over-reporting scales. We also hypothesized that scores derived from the testing-the-limits procedure would be less influenced by psychiatric symptoms and more strongly aligned with other measures of symptom over-reporting than the original M-FAST total score.

Method

Participants

Participants were recruited from the following: (1) a centralized database of veterans interested in participating in research at our center, (2) flyers posted at a US Department of Veterans Affairs (VA) medical center, and (3) group therapy sessions in a VA PTSD clinic, all using a standardized recruiting script. Eligible participants were veterans, at least 18 years old, who screened positive for current PTSD diagnosis over the telephone using the PTSD Checklist-5 (PCL-5; Weathers et al., 2014). Three hundred seventeen potential participants were screened via telephone. Two hundred sixty were eligible for the study, and 215 veterans agreed to participate. Of this group, six participants did not complete the protocol and were excluded from analyses, yielding a final sample of 209.

The sample comprised 175 male (83.7%) and 34 female veterans. Self-reported race and ethnicity were 65.07% white, 29.67% African American or black, 5.26% American Indian or Alaska Native, 2.87% unknown, 1.91% Asian, and 1.44% Native Hawaiian or Other Pacific Islander; 4.78% self-identified as Hispanic or Latino (categories were not exclusive; therefore, percentages sum to over 100). The mean age was 53.79 years (SD = 11.39) with a range of 21–75. The war eras for the 209 participants were as follows: Vietnam (39.23%), other (25.36%), Operation Iraqi Freedom/Operation Enduring Freedom/Operation New Dawn (22.97%), and Operation Desert Storm (11.48%). The prevalence of probable current PTSD diagnosis was 57.9% and of

lifetime PTSD diagnosis was 88.0% per the assessment and scoring approach described below.

Procedure

The study was approved by the VA Boston Healthcare System IRB, and all participants provided written informed consent at the start of the study. The protocol spanned approximately 3 h, and participants were compensated $75 at the end of the study. All self-report measures were completed via laptops. Interviews were administered by bachelor's-level psychology research technicians who received extensive training by the first author. All interviews were videotaped for reliability, consensus rating, and assessment fidelity purposes.

Measures

Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008; Tellegen & Ben-Porath, 2008) The MMPI-2-RF is a self-report inventory derived from the original MMPI-2 that measures personality, temperament, and psychopathology across 338 true-false items. The MMPI-2-RF includes 10 validity scales that evaluate response style for consistency, random responding, and over- and under-reporting of psychological and physical symptoms. The MMPI-2-RF also contains three Higher-Order scales that assess disordered thinking and aberrant perceptual experiences (Thought Dysfunction, THD), impulsive and rule-breaking behavior (Behavioral/ Externalizing Dysfunction, BXD), and mood and affect problems (Emotional/Internalizing Dysfunction, EID). Additionally, nine Restructured Clinical scales assess more unidimensional facets of these Higher-Order scales: RCd: Demoralization (generalized dissatisfaction, low self-esteem and agency); RC1: Somatic Complaints (physical complaints likely stemming from psychological origin); RC2: Low Positive Emotions (difficulty experiencing positive emotions, lack of social affiliation); RC3: Cynicism (distrust and negative opinions of others); RC4: Antisocial Behavior (impulsivity, unruliness); RC6: Ideas of Persecution (believing others pose a threat); RC7: Dysfunctional Negative Emotions (high levels of arousal and emotional intensity, anxiety, irritability); RC8: Aberrant Experiences (uncommon perceptual and cognitive experiences); and RC9: Hypomanic Activation (intense energy, aggression, grandiosity).

The MMPI-2-RF has undergone extensive reliability and validity assessment across inpatient, outpatient, veteran, and community samples. Repeatedly, its scales demonstrate expected associations with measures of both psychopathology and temperament, evidencing strong test-retest reliability and internal consistency (Tellegen & Ben-Porath, 2008). Because we were interested in invalid and malingered responses, we did not eliminate subjects from analyses based on MMPI-2-RF validity scales. Of those in our final sample of 209, there were 48 participants (22.97%) who produced invalid MMPI-2-RF profiles, as defined by Cannot Say (CNS) ≥ 18, VRIN T-score ≥ 80, TRIN T-score ≥ 80, or Fp-r T-score ≥ 100 (thresholds consistent with Arbisi et al., 2011).

Miller Forensic Assessment of Symptoms (M-FAST; Miller, 2001) The M-FAST is a short, 25-item forced-choice response style interview that measures possible suggestibility in symptom reporting and malingering. M-FAST items describe a range of unusual psychological phenomena that respondents endorse or deny. The interview lasts 5–10 min and is not thought to be influenced by literacy (Miller, 2005; see above for detailed description of this measure and its psychometric properties).

M-FAST: Testing-the-Limits Based on the concerns outlined above regarding the potential influence of factors other than malingering on M-FAST scores, we conducted a follow-up interview to the M-FAST that we developed for the purposes of this study in which we further queried participants about each endorsed item using a semi-structured interview and testing-the-limits approach. After completing the M-FAST per manual instructions, we returned to all endorsed items, one at a time, by stating: "A few moments ago you told me that you [insert endorsed item]. Tell me more about that." We developed item-specific follow-up questions to obtain additional details about the experience of each endorsed item and also instructed interviewers to probe for additional detail or clarification as needed, with an emphasis on use of open-ended questions. For example, if the participant endorsed an M-FAST item related to having unpleasant dreams that only co-occur with a specified physical symptom,[1] the interviewer would first ask the participant to provide more information (i.e., "tell me more about that") and then ask follow-up questions, such as: "Do you have these dreams … *only* when you have [insert physical symptom] or do you have them generally, not necessarily after [insert physical symptom]?" and "Why do you think you have these dreams only when [insert physical symptom]?"

During the course of this discussion, the interviewer recorded if the participant discontinued item endorsement and if so, the reason for this (e.g., the participant stated that they misheard or misunderstood the item when first presented). If the participant continued to endorse the item, the interviewer obtained enough information to determine if the participant was misunderstanding the item, responding in an idiosyncratic fashion, describing a plausible symptom (or some combination thereof), or malingering. The latter was often evident when the participant provided increasingly unconvincing or

[1] We are unable to provide more detailed descriptions of the M-FAST items, per the publisher of the M-FAST, due to copyright protections.

extreme support for their original item endorsement. This follow-up interview lasted approximately 5–10 min and was videotaped.

All videotapes were then reviewed by a consensus team consisting of at least two clinical psychologists and two psychology research assistants to make scoring decisions about each testing-the-limits item. In particular, the team determined if the participant still endorsed the item and if so, if the item should be scored as malingering. If the team determined that the participant no longer endorsed the item or that the response was not consistent with malingering, then the team also made ratings (that were not mutually-exclusive) concerning the reason for the changed score compared to the original M-FAST administration. Response options included the following: participant described plausible psychological or physical symptom, participant misunderstood or misheard original item, participant responded only to part of item, participant responded with an idiosyncratic interpretation of the item, or participant continued to endorse the item and appeared to be malingering. The guiding principle was that the evidence to change an original score based on information obtained in the testing-the-limits procedure had to be strong and that in the absence of strong evidence, the original score would be retained by default (e.g., that the participant should be scored in the direction of over-reporting). The consensus scores were analyzed in this study. The testing-the-limits interview was added later to the study; therefore, the sample size for analyses involving the testing-the-limits interview was $n = 176$. Cronbach's alphas for the testing-the-limits M-FAST scores are presented below.

Structured Interview of Malingered Symptomatology (SIMS; Widows & Smith, 2005; Smith & Burger, 1997) The SIMS is a 75-item, true/false self-report inventory of malingered symptoms. The original validation sample included college students attempting to feign severe psychopathology (Smith & Burger, 1997). The SIMS takes approximately 15 min to complete and has demonstrated associations with malingering as determined via clinical interview (Lewis, Simcox, & Berry, 2002; Wisdom, Callahan, & Shaw, 2010). The SIMS was also found to correctly identify individuals instructed to feign PTSD symptoms (Merten, Lorenz, & Schlatow, 2010) and brain injury (Jelicic, Ceunen, Peters, & Merckelbach, 2011). The SIMS was added later to the protocol, yielding a sample size for SIMS-related analyses of $n = 173$. Cronbach's alpha coefficient for the 75 items within this sample was $\alpha = 0.85$. The measure includes five subscales that assess different types of feigned symptoms, including those related to psychosis, neurological impairment, amnestic disorders, low intelligence, and affective disorders. Cronbach's alpha coefficients for these scales in this sample were α = 0.77, 0.72, 0.73, 0.21, and 0.38, respectively. Total score ≥ 14 was used as the threshold for potential malingering in this study (Widows & Smith, 2005), though we also examined other cut-points identified in the SIMS manual (scores of 13+ and 16+; Widows & Smith, 2005) and a much higher one (44+) identified by Rogers, Robinson, and Gillard (2014).

Privacy Expectations (PE) Scale To assess participants' understanding of the privacy and confidentiality of their research data within this study, we developed an 11-item self-report questionnaire. The PE Scale assesses the extent to which participants believed their data could be used for purposes other than those outlined in the oral and written informed consent procedures (i.e., for disability appeals, to inform their clinical care, or for legal cases). Each item was assessed on a 5-point Likert scale (1 = "strongly disagree" to 5 = "strongly agree") indicating the extent to which each participant expected that their data would be shared or used for purposes outside of the research. Cronbach's alpha coefficient for the 11 items within our sample was $\alpha = 0.83$. Mean scores were analyzed, with higher scores indicating that the participant believed the study data would be used for purposes outside of the research.

Trauma Assessment from the National Stressful Events Survey (NSES; Kilpatrick, Resnick, Baber, Guille, & Gros, 2011) The NSES was developed to assess trauma exposure and *DSM-5* PTSD symptoms. The trauma component of the NSES was administered as an interview to examine exposure to traumatic events pre-military (natural or man-made disaster, terrorist attack, physical or sexual abuse, serious accidents or fire, sudden death of a close relative or friend, witnessing dead bodies, other stressful events causing injury or fear of being killed or injured), military (combat or its aftermath, military sexual trauma (MST), military training accident, or other stressful events related to military service), and post-military (same events listed for pre-military). Participants were asked to identify their "worst" traumatic experience. The NSES has been used for web-based assessment of *DSM-5* PTSD in community and veteran samples (Miller et al., 2013) and has shown good internal consistency ($\alpha = 0.94$) and strong correlations with *DSM-5*-defined PTSD severity.

DSM-5 PTSD Symptom Assessment (Adapted from Weathers et al., 2014 and Kilpatrick et al., 2011; See Also Miller et al., 2013) In order to assess probable current PTSD, we administered items (via interview) from the PTSD Checklist-5 (PCL-5; Weathers et al., 2014) anchored to the participant's self-identified "worst" traumatic experience as reported on the NSES (Kilpatrick et al., 2011). The PCL-5 is a 20-item self-report measure of PTSD symptoms that are aligned with the *DSM-5* PTSD criteria. Each item is rated on a 5-point scale to denote the extent to which the participant has been distressed by the symptom in the past month (0 = "not at all" to 4 = "severe"). The factor structure aligns with the PTSD symptom clusters outlined in the *DSM-5*; in addition, the PCL-5 has performed well with respect to internal consistency, test-

retest reliability, and convergent and discriminant validity (Bovin et al., 2016; Blevins, Weathers, Davis, Witte, & Domino, 2015).

We adapted the response options to also include evaluation of lifetime symptomatology in order to better ensure that reports of current symptoms were reflective of only the past month (Kilpatrick et al., 2011; Miller et al., 2013). Using the standard PCL-5 item phrasing, we first asked if the participant had ever experienced a given symptom, and if so, whether the symptom was present in the past month. Only if a given past month symptom was endorsed did we then ask participants to rate the current severity of that symptom on the standard PCL-5 response scale. Current probable PTSD diagnostic status was defined by symptom presence ratings of "moderate" severity or greater (e.g., a score ≥ 2) on the requisite number of symptoms in each *DSM-5* defined cluster (Bovin et al., 2016). Cronbach's alpha coefficient for past month symptom severity ratings was $\alpha = 0.87$.

The Medical Symptom Validity Test (MSVT; Green, 2004) The MSVT is a brief, computerized self-report and oral assessment of verbal memory and response consistency. Individuals who do not "pass" this memory test are identified as likely to be feigning cognitive problems or to be putting forth poor effort towards completing the study protocol, thus raising questions about the validity of their self-report responses. Prior research with veterans suggests that just under 10% of participants "fail" this test (Clark, Amick, Fortier, Milberg, & McGlinchey, 2014), and that the MSVT has excellent sensitivity and specificity (Carone, 2009).

Data Analysis

We first examined the frequency of endorsement of each M-FAST total score and of each individual M-FAST item (in the full sample and separated by probable current PTSD diagnosis). We did the same for scores derived from the testing-the-limits procedure, and we also quantified clinician determinations concerning the reasons why an item was no longer rated as malingering (e.g., the participant misunderstood the question, the clinician consensus determined the participant reported a plausible psychological symptom; see above). We also estimated the prevalence of malingering per the M-FAST 6+ total score rule and compared this to classification of invalid (see above definition) MMPI-2-RF and SIMS profiles using χ^2. We also examined how this classification differed by probable PTSD diagnosis, sex, and minority status and then examined how alternate cut-scores (identified above) affected the prevalence of symptom over-reporting determinations. We then calculated Cronbach's alpha and average interitem correlation (AIC) to determine internal consistency of the scale overall and of each subscale (for scales with more than one item). AIC has a recommended range of 0.15 to 0.50 (Clark & Watson, 1995). We conducted an exploratory factor analysis (EFA) with the robust maximum likelihood estimator (MLR) and geomin rotation to examine the underlying factor structure of the measure and explicitly compared the one-factor model described previously (Miller, 2001; Vitacco et al., 2008) with multi-factor models, with the maximum factors determined by the scree plot. We then conducted item-response theory (IRT) analyses to determine the relationship between each item and the malingering construct overall and the information provided by each item. IRT examined both item characteristic curves (ICC) and item information curves (IIC). The ICC shows the probability of endorsing an item given the level of the trait (malingering) present, and the IIC can be conceptualized as depicting the range of the latent trait at which the information of a given item is maximized, showing the precision of the item. Poorly performing items were eliminated. We then conducted a second EFA of the retained items to determine if this improved the fit and interpretation of underlying factors.

We next compared the pattern of correlations between M-FAST total scores, all-measure factor analysis-based summary scores (as determined by the EFA of all items), retained-item factor-analysis based summary scores (as determined by the EFA of IRT-retained items only), and revised M-FAST total scores from the testing-the-limits interview in association with each other and with the following external correlates: demographic variables (including those pertaining to compensation seeking); MMPI-2-RF Validity, Higher-Order, and Restructured Clinical scales; the SIMS; the MSVT; the PE Scale; and PTSD symptom severity. Finally, to examine the relative association between over-reporting versus psychopathology with M-FAST, we conducted regressions with the best MMPI-2-RF over-reporting scale (Fp-r in PTSD samples, Marion et al., 2011; Sellbom & Bagby, 2010) and the three MMPI-2-RF Higher-Order scales as predictors of M-FAST scores. All latent variable analyses were conducted in Mplus v. 8.0 (Muthén & Muthén, 1998-2017) and their relative fit was evaluated using Akaike information criterion, Bayesian information criterion (BIC), and the robust nested χ^2 test. All other analyses were conducted in SPSS v. 25.

Results

Total Scale and Item-Level Endorsement

Original M-FAST total scores ranged from 0 to 17 with a mean of 3.25 (*SD* 3.16; Table 1). Individual M-FAST item endorsement ranged from 1.9% (item 6) to 51.2% (item 2; Table 2) with items 2, 4, 7, 9, 11, 14, 15, 17, and 20–24 more common among probable PTSD cases than those without the diagnosis (all $p < .05$; Table 2). Using the recommended cut score of 6+ items, 17.2% of the sample ($n = 36$) met criteria for malingering according to the original M-FAST total score, and

Table 1 Descriptive statistics and internal consistency estimates of M-FAST total and subscale scores

Scale	Original scoring				Testing-the-Limits scoring			
	Full sample *M* (*SD*)	PTSD+ *M* (*SD*)	PTSD− *M* (*SD*)	Full sample α (mean AIC)	Full sample *M* (*SD*)	PTSD+ *M* (*SD*)	PTSD− *M* (*SD*)	Full sample α (mean AIC)
Total score	3.25 (3.16)	4.21 (3.60)***	1.93 (1.73)	0.78 (0.13)	1.11 (1.91)	1.53 (2.28)***	0.47 (0.86)	0.74 (0.11)
RO	0.34 (0.54)	0.43 (0.58)*	0.22 (0.47)	0.00 (0.01)	0.14 (0.36)	0.17 (0.40)	0.09 (0.28)	0.008 (0.00)
ES	0.98 (0.89)	1.22 (0.95)***	0.64 (0.65)	0.19 (0.07)	0.24 (0.55)	0.34 (0.65)*	0.10 (0.30)	0.28 (0.07)
RC	0.77 (1.18)	1.01 (1.38)***	0.43 (0.72)	0.60 (0.17)	0.30 (0.65)	0.40 (0.77)*	0.14 (0.39)	0.40 (0.09)
UH	0.52 (0.84)	0.69 (0.96)**	0.28 (0.57)	0.48 (0.14)	0.26 (0.64)	0.39 (0.77)**	0.07 (0.31)	0.52 (0.15)
USC	0.14 (0.35)	0.21 (0.41)**	0.05 (0.21)	N/A	0.07 (0.25)	0.11 (0.31)*	0.01 (0.12)	N/A
NI	0.44 (0.50)	0.52 (0.50)**	0.26 (0.44)	N/A	0.04 (0.20)	0.04 (0.19)	0.4 (0.20)	N/A
Sugg	0.08 (0.27)	0.09 (0.29)	0.06 (0.23)	N/A	0.03 (0.18)	0.04 (0.19)	0.03 (0.17)	N/A

$n = 209$ for original administration and $n = 176$ for testing-the-limits administration

M-FAST Miller Forensic Assessment of Symptoms Test, *PTSD* posttraumatic stress disorder, *RO* Reported vs. Observed, *ES* Extreme Symptomatology, *RC* Rare Combinations, *UH* Unusual Hallucinations, *USC* Unusual Symptom Course, *NI* Negative Image, *Sugg* Suggestibility, *N/A* not applicable (due to scale consisting of 1 item)

*$p < .05$; **$p < .01$; ***$p < .001$

Table 2 M-FAST item-level endorsement

Item no.	Original scoring			Testing-the-limits scoring		
	Full sample (%)	PTSD+ (%)	PTSD− (%)	Full sample (%)	PTSD+ (%)	PTSD− (%)
1	19.1	19.8	18.2	6.7	8.6	7.0
2	51.2	67.8***	28.4	3.8	5.7	2.8
3	12.9	14.9	10.2	1.9	1.9	2.8
4	2.9	5.0*	0.0	1.0	1.9	0.0
5	19.6	17.4	22.7	2.9	3.8	2.8
6	1.9	1.7	2.3	1.0	1.0	1.4
7	8.6	13.2**	2.3	2.4	3.8	1.4
8	5.7	7.4	3.4	2.4	4.8	0.0
9	7.2	10.7*	2.3	3.8	6.7	1.4
10	4.8	5.0	4.5	2.4	3.8	1.4
11	6.2	9.9**	1.1	2.4	4.8	0.0
12	7.7	9.1	5.7	2.4	2.9	2.8
13	1.0	1.7	0.0	15.8	0.0	0.0
14	13.9	20.7**	4.5	5.7	10.5*	1.4
15	11.0	17.4**	2.3	8.1	14.3*	2.8
16	12.0	15.7	6.8	5.7	9.5	2.8
17	16.7	24.0**	6.8	6.7	11.4*	2.8
18	9.6	12.4	5.7	3.8	5.7	2.8
19	4.3	5.8	2.3	1.0	1.9	0.0
20	22.5	28.9**	13.6	8.6	16.2**	1.4
21	17.2	22.3*	10.2	7.7	13.3*	2.8
22	12.9	17.4*	6.8	1.9	3.8	0.0
23	43.5	56.2***	26.1	3.3	3.8	4.2
24	4.8	7.4*	1.1	2.4	4.8	0.0
25	7.7	9.1	5.7	2.9	3.8	2.8

$n = 209$ for original scoring and $n = 176$ for testing-the-limits scoring

M-FAST Miller Forensic Assessment of Symptoms Test, *PTSD* posttraumatic stress disorder

Table 3 Effect of differing cut scores on identification of symptom over-reporting

Cut scores	Overall		PTSD+		PTSD−		χ^2
	%	*n*	%	*n*	%	*n*	
M-FAST							
6	17.22	36	26.45	32	4.55	4	17.14***
8	11.48	24	19.01	23	1.14	1	16.01***
16	0.48	1	0.83	1	0.00	0	0.73
SIMS							
13	70.18	120	86.54	90	44.78	30	33.96***
14	66.67	114	82.69	86	41.79	28	30.68***
16	59.65	102	74.04	77	37.31	25	22.83***
44	2.34	4	3.85	4	0.00	0	2.64

$n = 209$ for analyses of the M-FAST; $n = 171$ for analyses of the SIMS. The M-FAST cut-score of 6 is recommended in the manual while scores of 8 and 16 have been described in the literature (see introduction). The SIMS cut score of 14 is recommended in the measure's manual, and differential sensitivity and specificity at cut scores of 13 and 16 are also described in the manual; a cut score of 44 has been evaluated in prior published research (see introduction)

M-FAST Miller Forensic Assessment of Symptoms Test (Miller, 2001), *SIMS* Structured Interview of Malingered Symptomatology (SIMS; Widows & Smith, 2005; Smith & Burger, 1997)

***$p < .001$

this was more prevalent among probable current PTSD cases (26.45%) compared to controls (4.55%): χ^2 (1, $n = 209$) = 17.14, $p < .001$ (Table 3). The prevalence of symptom over-reporting did not appreciably change in the probable PTSD cases when a cut-score of 8+ was applied, but did drop considerably at a score of 16+ (Table 3). Using the MMPI-2-RF, 23.0% ($n = 48$) were classified as over-reporting using the validity score rules described above (31.1% among those with probable current PTSD diagnosis versus 12.8% among those without, χ^2 [1, $n = 205$] = 9.33, $p = .002$), and this was more common among those identified as malingering per the M-FAST (61.1%) versus those under the threshold (15.4%): χ^2 (1, $n = 205$) = 34.61, $p < .001$. In the subset of subjects (n = 171) who completed the SIMS, 66.67% were identified as potentially malingering using the recommended cut-score of 14+ (82.69% with probable current PTSD versus 41.79% among those without, χ^2 [1, $n = 171$] = 30.68, $p < .001$). The prevalence of symptom over-reporting classifications on the SIMS among the probable PTSD sample at alternate cut-scores did not change substantially until a very high cut-point was applied (Table 3). All participants who screened positive for malingering on the M-FAST were also classified as such on the SIMS, while 60.1% of those who did not meet the M-FAST threshold were identified as malingering on the SIMS: χ^2 (1, $n = 171$) = 16.74, $p < .001$. There were 17 participants (8.1%) who scored above the recommended thresholds on all three measures.

Using the recommended M-FAST cut-score of 6+, there were no differences in over-reporting classifications as a function of race (14.41% among those who identified as white vs. 21.69% among those who identified as a racial minority, χ^2 [1, $n = 201$] = 1.80, $p = .180$). There was also no difference by ethnicity (16.67% among those who identified as non-Hispanic vs. 10.00% among those who identified as Hispanic, χ^2 [1, $n = 196$] = .31, $p = .578$. The prevalence of over-reporting classification also did not differ by sex (12.50% among women vs. 17.37% among men, χ^2 [1, n = 199] = .46, $p = .498$).

Testing-the-limits M-FAST total scores ranged from 0 to 13 with a mean of 1.11 (*SD* 1.81; Table 1) and individual item endorsement ranged from 1.0 to 15.8% (Table 2). There was substantial item-level variability in the extent to which item scores changed during the testing-the-limits procedure (Table 4). For example, items 6 and 13 (which are part of the Unusual Hallucinations and Extreme Symptomatology subscales, respectively) were rarely changed (0.6% of the sample each) whereas item 2 (Extreme Symptomatology subscale) was commonly changed to a not-over-reporting determination (45.5% of the sample), which was overwhelmingly due to consensus determination that the veteran reported a plausible psychological symptom related to low mood (92.5% of changed ratings). Some items (nos. 3–8, 10–16, 19, 21, 25) were more prone to the participant no longer endorsing the item when asked to further explain and discuss the item, and this was commonly a function of the participant mishearing or misunderstanding the item or to idiosyncratic item interpretation. In contrast, other items (nos. 1–2, 9, 17–18, 20, 22–23) were more prone to being changed as a function of consensus determinations, despite the participant continuing to endorse the item. This was typically due to the team determining that the symptom was a plausible psychological symptom (Table 4).

Internal Consistency

Coefficient alpha for the M-FAST total scale was $\alpha = 0.78$ and mean AIC was 0.13 (Table 1). Internal consistency for the M-FAST subscales was also examined and revealed several unacceptably low values: $\alpha = 0.00$ for Reported vs. Observed, $\alpha = 0.19$ for Extreme Symptomatology, and $\alpha = 0.48$, for Unusual Hallucinations, while the estimate for Rare Combinations was relatively stronger at $\alpha = 0.60$ (Table 1). Estimates of internal consistency for scale scores derived from the testing-the-limits protocol were also generally poor (Table 1).

Table 4 M-FAST testing-the-limits changes to item scoring

Item no.	Score changed % (*n*)	Participant no longer endorsed % (*n*)	Consensus determination % (*n*)	Misheard or misunderstood % (*n*)	Responded to part of item % (*n*)	Idiosyncratic interpretation % (*n*)	Plausible symptom % (*n*)
1	10.2 (18)	27.8 (5)	72.2 (13)	11.1 (2)	11.1 (2)	11.1 (2)	83.3 (15)
2	45.5 (80)	3.8 (3)	96.3 (77)	1.3 (1)	3.8 (3)	3.8 (3)	92.5 (74)
3	10.8 (19)	57.9 (11)	42.1 (8)	31.6 (6)	26.3 (5)	36.8 (7)	42.1 (8)
4	2.3 (4)	100 (4)	0.0	25 (1)	50 (2)	50 (2)	0.0
5	16.5 (29)	62.1 (18)	37.9 (11)	48.3 (14)	25 (7)	50 (14)	3.6 (1)
6	0.6 (1)	100 (1)	0.0	0.0	100 (1)	0.0	0.0
7	5.1 (9)	55.6 (5)	44.4 (4)	44.4 (4)	44.4 (4)	22.2 (2)	55.6 (5)
8	2.8 (5)	100 (5)	0.0	40 (2)	60 (3)	20 (1)	0.0
9	3.4 (6)	33.3 (2)	66.7 (4)	16.7 (1)	0.0	66.7 (4)	16.7 (1)
10	2.3 (4)	100 (4)	0.0	25 (1)	50 (2)	75 (3)	0.0
11	3.4 (6)	83.3 (5)	16.7 (1)	16.7 (1)	83.3 (5)	16.7 (1)	16.7 (1)
12	5.7 (10)	80 (8)	20 (2)	20 (2)	70 (7)	20 (2)	10 (1)
13	0.6 (1)	100 (1)	0.0	0.0	100 (1)	100 (1)	0.0
14	7.4 (13)	61.5 (8)	38.5 (5)	15.4 (2)	23.1 (3)	53.8 (7)	23.1 (3)
15	1.1 (2)	100 (2)	0.0	50 (1)	50 (1)	50 (1)	0.0
16	5.1 (9)	66.7 (6)	33.3 (3)	44.4 (4)	44.4 (4)	33.3 (3)	22.2 (2)
17	8.0 (14)	14.3 (2)	85.7 (12)	14.3 (2)	14.3 (2)	7.1 (1)	78.6 (11)
18	5.1 (9)	33.3 (3)	66.7 (6)	22.2 (2)	11.1 (1)	44.4 (4)	55.6 (5)
19	4.0 (7)	57.1 (4)	42.9 (3)	0.0	0.0	100 (7)	14.3 (1)
20	9.7 (17)	23.5 (4)	76.5 (13)	0.0	23.5 (4)	5.9 (1)	70.6 (12)
21	8.0 (14)	78.6 (11)	21.4 (3)	21.4 (3)	7.1 (1)	78.6 (11)	7.1 (1)
22	11.9 (21)	38.1 (8)	61.9 (13)	38.1 (8)	14.3 (3)	42.9 (9)	33.3 (7)
23	37.5 (66)	12.1 (8)	87.9 (58)	0.0	9 (6)	16.4 (11)	86.6 (58)
24	1.1 (2)	50 (1)	50 (1)	0.0	0.0	100 (2)	0.0
25	2.8 (5)	60 (3)	40 (2)	40 (2)	20 (1)	0.0	60 (3)
Mean	8.4%	60%	52.7%	27.7%	34.2%	43.7%	40.7%

$n = 176$. The denominator for the first column is 176 (total number of subjects who completed testing the limits). The denominator for all other columns is the number of times a score was changed for that item (i.e., column 1). Whether the participant no longer endorsed the item upon follow-up questioning or whether it was still endorsed but the consensus was that the symptom did not represent malingering was a mutually exclusive determination. However, detailed explanations (last 4 columns) of why the symptom score was changed were not mutually exclusive and therefore sum to more than 100% for a given item

EFA of all Items

We estimated EFA models with one to four factors and examined their relative fit. The scree plot (Fig. S1) supported the existence of two factors, at most; thus, the one- and two-factor solutions are reported. All but two items loaded significantly ($p < .05$) on the latent variable in the one-factor model (Table S1). Items 1 and 5 had no association with the latent variable (βs = 0.00 and 0.02, respectively). Item 12 also loaded weakly ($\beta = 0.34$) on the latent variable and several additional items (2, 3, 6, 10) loaded at $\beta < 0.50$. The two-factor model did not yield clearly better fit than the one-factor model, further supporting the decision not to pursue additional factors. Specifically, while the robust χ^2 comparison supported the parent two-factor model over the more restricted one-factor model, the BIC supported the one-factor model and the pattern of factor loadings in the two-factor solution included six items without a significant loading on either factor (Tables S1-S2).

Item Response Theory

IRT-based ICC and IIC curves are shown in Figs. 1a–e and 2a–e, respectively. The *x*-axis is theta, or the level of the latent trait, and the *y*-axis depends on the type of curve shown: In ICCs, the *y*-axis is the probability of endorsing that item given the level of the latent trait; in IICs, the *y*-axis is the information provided by the item at different levels of the latent trait. The curves presented for

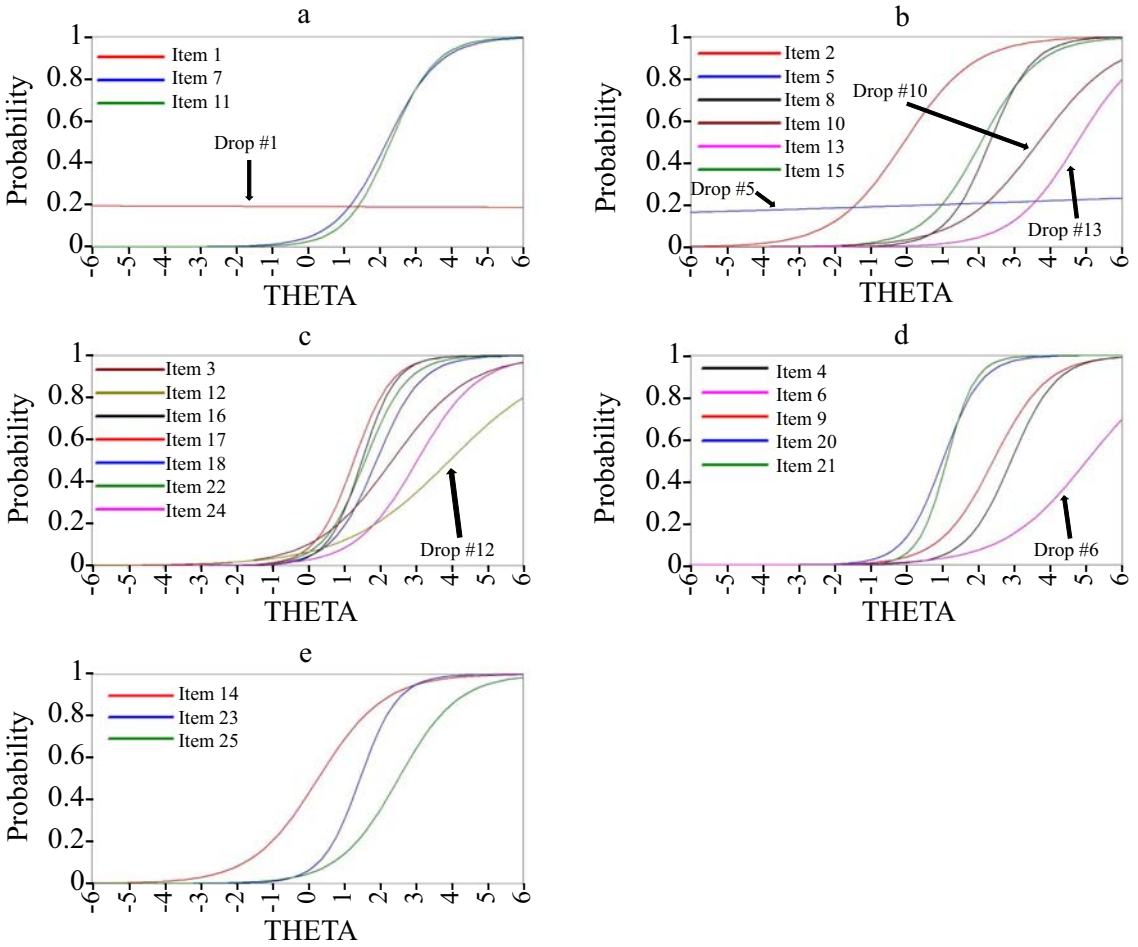

Fig. 1 Item characteristic curves for M-FAST items; *x*-axis is theta or level of the latent construct malingering, and *y*-axis is the information provided by the item. Analyses were based on the entire M-FAST, but curves are divided into separate panels by the M-FAST subscales to aid interpretation. **a** Items in the Reported versus Observed subscale. **b** Items in the Extreme Symptomatology subscale. **c** Items from the Rare Combinations subscale. **d** Unusual Hallucinations subscale items. **e** Subscales constituted by a single item (i.e., Unusual Symptom Course, Negative Image, and Suggestibility). Items that were deemed as poorly performing in the IRT analyses are indicated by the caption "Drop #X"

ICCs are a logistic function, allowing comparisons among all items. A curve indicating good ICC is tall by ranging across probability on the y-axis, and steep, indicating that the presence of a small range of theta quickly increases the probability of endorsing the item (i.e., that individuals with above average scores on the latent trait are likely to endorse the item). Items showing good IIC curves have peaked distributions, with the taller the peak corresponding to more information. A flat distribution indicates similar levels of information across levels of theta. Short and flat curves indicate low precision and thus low reliability.

IRT analyses were conducted for the entire M-FAST scale but are depicted here by subscale simply to make figures easier to read. For Reported vs. Observed items, item 1 provided little information across all levels of theta, whereas items 11 and 7 displayed good IICs (Fig. 1a). Similarly, the ICC (Fig. 2a) showed essentially flat and short curves across all levels of theta for item 1, whereas the ICC for items 7 and 11 were acceptable; thus, we eliminated item 1. For Extreme Symptomatology items, the IIC for 5, 10, and 13 were problematic (Fig. 1b). The IIC for item 5 was flat and provided no information across all levels of theta. Items 10 and 13 provided little information, and item 13 yielded information only at extremely high levels of theta. The ICCs (Fig. 2b) were consistent in that the item 5 curve was essentially flat across all levels of theta and anchored to a low probability level (indicating that it did not discriminate between malingering and not malingering). Items 10 and 13 did not exhibit sharp curves, indicating weak discrimination. More importantly, these items were shifted high on the *x*-axis, indicating a high degree of difficulty: only those very high on the latent construct would endorse the items, but these items still did not discriminate well between those high and low on malingering. Based on these findings, items 5, 10, and 13 were dropped.

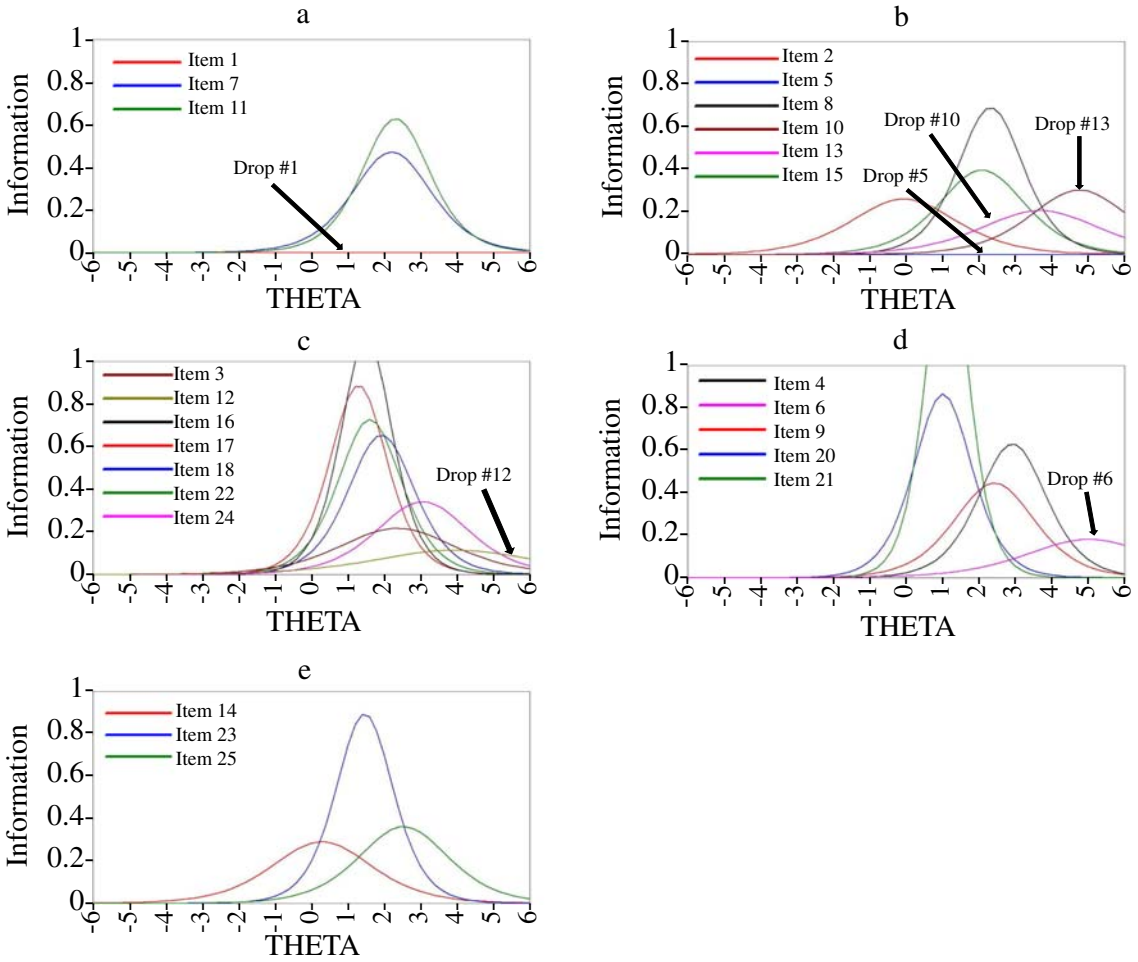

Fig. 2 Item information curves for M-FAST items; *x*-axis is theta or level of the latent construct malingering, and *y*-axis is the probability of item endorsement. Analyses were based on the entire M-FAST, but curves are divided into separate panels by the M-FAST subscales to aid interpretation. **a** Items in the Reported versus Observed subscale. **b** Items in the Extreme Symptomatology subscale. **c** Items from the Rare Combinations subscale. **d** Unusual Hallucinations subscale items. **e** Subscales constituted by a single item (i.e., Unusual Symptom Course, Negative Image, and Suggestibility). Items that were deemed as poorly performing in the IRT analyses are indicated by the caption "Drop #X"

Most Rare Combinations items performed well in the IRT analyses, with the exception of items 3 and 12. These items showed flat IICs (Fig. 1c) and ICCs (Fig. 2c) that were shifted to the right but with less discrimination between endorsement and lack of endorsement of the item; thus, these items were eliminated. Items in the Unusual Hallucinations scale tended to perform well, except for item 6, which evidenced a flat IIC (Fig. 1d) that showed an extreme right shift, indicating not only that this item was imprecise, but also that its highest level of precision was only for those at the extremely high end of the construct. This item also evidenced poor discrimination on the ICC (Fig. 2d) and was eliminated. IRT results for the items from the single-item subscales are grouped together for presentation purposes only (Figs. 1e and 2e). All three items displayed fair to good IIC and ICC curves and were retained. In summary, seven items displayed poor psychometric properties in this sample (nos. 1, 3, 5, 6, 10, 12, 13) according to IRT analyses. These were also the weakest loadings in the all-item EFA and were therefore eliminated from the follow-up EFA in which we

Table 5 Intercorrelations among M-FAST-based total scores

Scale	1	2	3
1. Original			
2. Reduced item set	0.97***		
3. TTL original	0.80***	0.79***	
4. Reduced item set TTL	0.78***	0.81***	0.95***

$n = 209$ for correlations based on traditional scale administration and $n = 176$ for correlations involving the testing-the-limits scores

M-FAST Miller Forensic Assessment of Symptoms Test, *TTL* testing-the-limits

***$p < .001$

tested a refined item set that could potentially be a stronger index of malingering.

EFA of Revised Item-Set Based on Initial EFA and IRT

The EFA (1–4 factors) of the revised item set which eliminated seven poorly performing items resulted in a scree plot (Fig. S2) that again suggested a maximum of two factors and fit statistics and patterns of factor loadings that showed the clear superiority of the one-factor solution (Tables S1-S2). Specifically, the one-factor solution yielded a lower BIC value than the two-factor model and the robust χ^2 difference test suggested that the one-factor model did not damage fit relative to the parent two-factor model. Further, in the one-factor model, all items loaded significantly ($p < .05$) on the latent variable, while in the two-factor model, there were no significant loadings (Table S1). Thus, the one-factor model was determined to be the best fitting model.

Correlations Among M-FAST Scores

Based on the EFA results supporting a single factor, we next evaluated patterns of correlations among original M-FAST total scores, reduced item set M-FAST total scores, original testing-the-limits M-FAST total scores, and reduced item set testing-the-limits scores (Table 5). In general, all the scales were highly correlated with each other (smallest $r = .78$), both across scales that differed by item inclusion/exclusion and across the original and testing-the-limits protocols.

Correlations Between M-FAST Scores and External Measures

We next compared the pattern of correlations across the four M-FAST-based scales and a variety of external correlates (Table 6). We examined the original M-FAST total scores, the total score with poorly performing items excluded, the original testing-the-limits total score, and the testing-the-limits total score with the same items eliminated (for comparison purposes). To test for the statistical significance of the differences between correlations, we used the Williams modification of the Hoteling test for two correlations involving a common variable (Kenny, 1987). The highest correlations within each row (i.e., the M-FAST scale with the strongest correlation for each variable) are indicated by the subscript "a," and ascending subscript letters denote statistically weaker correlations differentiated from each other. In general, correlations tended to show similar patterns of association across all four M-FAST scales, with a tendency for the testing-the-limit scores to be significantly lower. Basic demographic characteristics (age, sex, racial or ethnic minority) were not significantly correlated with any of the M-FAST scales. Receiving compensation for PTSD from the VA was weakly (and equivalently) associated with M-FAST scores, while being in mental health treatment was negatively associated with original M-FAST scores but not with testing-the-limits scores. Although the testing-the-limits scores were significantly less strongly associated with psychiatric symptoms (on both the MMPI-2-RF and the PTSD measure) than were the original and reduced item set M-FAST scores (consistent with hypotheses), they were also significantly less strongly related to MMPI-2-RF and SIMS-based measures of over-reporting (inconsistent with hypotheses). Of note, F-r was the MMPI-2-RF validity scale with the strongest associations with each M-FAST scale, although it was not statistically different in some cases from Fp-r (three out of four M-FAST scales), F-s (two out of four M-FAST scales), and RBS (two out of four M-FAST scales). Of the Higher-Order MMPI-2-RF scales, THD evidenced the strongest associations with each M-FAST scale (zs = 4.06 to 9.84; $p < .05$), which were stronger than correlations between the M-FAST scales and PTSD severity (zs = 2.92 to 5.12; $p < .05$) and MMPI-2-RF validity scales, with the sole exception of equivalence with F-r for the total score with the poorly performing items excluded (zs = 2.01 to 10.08; $p < .05$). The SIMS Psychosis subscale was the strongest SIMS scale in association with the M-FAST scales, although it was not significantly higher than the Neurological subscale for both testing-the-limits scores, and was additionally not higher than the Amnestic subscale with the testing-the-limits reduced item set. Of note, all M-FAST scales were significantly correlated with MMPI-2-RF TRIN scales, suggesting the possibility of a "yeah-saying" bias in item endorsement. There were no significant associations between any M-FAST score and an established measure (the MSVT) of poor effort on neurocognitive tasks and there was only a weak association between the original M-FAST total score and the expectation that data would be used for personal gain (as indicated by the PE Scale).

Relative Associations between Over-Reporting, Psychopathology, and M-FAST

Table 7 reveals the results of regressions in which the MMPI-2-RF over-reporting scale Fp-r, the three Higher-Order scales, and current PTSD symptom severity were included as predictors of the reduced item set M-FAST total scores and the equivalent total score from the testing-the-limits procedure. These analyses revealed that MMPI-2-RF-defined over-reporting was no longer significantly associated with either M-FAST score when psychopathology was included in the model. In addition, THD evidenced the strongest association with M-FAST scores and was unchanged across the reduced item set score from the original versus the testing-the-limits protocols. In contrast, both EID and BXD were associated with the original M-FAST total score (with the latter

Table 6 Correlations between M-FAST-based scores and external correlates

Measure	Original	Reduced item set	TTL original	Reduced item set TTL
Demographics				
Age	− 0.03	− 0.03	− 0.01	0.01
Sex	− 0.03	− 0.02	− 0.06	− 0.08
Minority	0.11	0.08	0.12	0.10
PTSD service connection	0.19_{a}^{*}	0.18_{a}^{*}	0.19_{a}^{*}	0.15
SSDI for PTSD	0.04	0.06	0.01	0.04
Legal for PTSD	0.11	0.13	0.12	0.13
In mental health tx	-0.17_{a}^{*}	-0.18_{a}^{*}	− 0.08	− 0.12
MMPI-2-RF				
Validity Scales				
VRIN	0.03	− 0.01	− 0.06	− 0.08
TRIN	0.19_{a}^{**}	0.18_{a}^{*}	0.18_{a}^{*}	0.16_{a}^{*}
F-r	0.66_{b}^{***}	0.71_{a}^{***}	0.43_{c}^{***}	0.45_{c}^{***}
Fp-r	0.59_{a}^{***}	0.61_{a}^{***}	0.36_{b}^{***}	0.37_{b}^{***}
F-s	0.56_{b***}	0.60_{a}^{***}	0.43_{d}^{***}	0.47_{c}^{***}
FBSr	0.40_{b}^{***}	0.45_{a}^{***}	0.26_{c}^{**}	0.30_{c}^{***}
RBS	0.53_{b}^{***}	0.58_{a}^{***}	0.35_{c}^{***}	0.38_{c}^{***}
Lr	− 0.05	− 0.09	− 0.05	− 0.07
Kr	-0.50_{a}^{***}	-0.43_{b}^{***}	-0.27_{c}^{***}	-0.26_{c}^{**}
Higher-Order Scales				
EID	0.46_{b}^{***}	0.53_{a}^{***}	0.27_{c}^{***}	0.29_{c}^{***}
THD	0.74_{a}^{***}	0.73_{a}^{***}	0.57_{b}^{***}	0.56_{b}^{***}
BXD	0.22_{a}^{**}	0.23_{a}^{**}	0.13	0.13
Restructured Clinical Scales				
RCd	0.46_{b}^{***}	0.52_{a}^{***}	0.29_{c}^{***}	0.29_{c}^{***}
RC1	0.48_{b}^{***}	0.53_{a}^{***}	0.38_{c}^{***}	$0.41_{b,c}^{***}$
RC2	0.29_{b}^{***}	0.37_{a}^{***}	0.15	0.18_{c}^{*}
RC3	0.33_{a}^{***}	0.33_{a}^{***}	$0.26_{a,b}^{**}$	0.24_{b}^{**}
RC4	0.27_{a}^{***}	0.27_{a}^{***}	0.15_{b}^{*}	0.14
RC6	0.68_{a}^{***}	0.66_{a}^{***}	0.50_{b}^{***}	0.49_{b}^{***}
RC7	0.51_{a}^{***}	0.53_{a}^{***}	0.34_{b}^{***}	0.33_{b}^{***}
RC8	0.67_{a}^{***}	0.67_{a}^{***}	0.52_{b}^{***}	0.52_{b}^{***}
RC9	0.31_{a}^{***}	0.31_{a}^{***}	0.26_{a}^{**}	0.25_{a}^{**}
SIMS				
Total score	0.70_{b}^{***}	0.73_{a}^{***}	0.53_{c}^{***}	0.53_{c}^{***}
Neurological	0.60_{a}^{***}	0.62_{a}^{***}	0.47_{b}^{***}	0.47_{b}^{***}
Affective	0.43_{b}^{***}	0.47_{a}^{***}	0.25_{c}^{**}	0.27_{c}^{**}
Psychosis	0.74_{a}^{***}	0.74_{a}^{***}	0.56_{b}^{***}	0.55_{b}^{***}
Low Intelligence	0.31_{a}^{***}	0.32_{a}^{***}	0.26_{a}^{**}	0.25_{a}^{**}
Amnestic	0.56_{b}^{***}	0.61_{a}^{***}	0.43_{d}^{***}	0.47_{c}^{***}
MSVT fail	− 0.09	− 0.11	− 0.05	− 0.09
Total trauma exposure	0.15^{*}	0.12	0.14	0.11
Current PTSD severity	0.51_{b}^{***}	0.54_{a}^{***}	0.39_{c}^{***}	0.40_{c}^{***}
Privacy expectations scale	0.15^{*}	0.11	0.10	0.07

Lowercase letters denote a statistically significant difference in strength between correlations within the same row: "a" denotes a correlation is the strongest in its row, "b" denotes it is the second strongest, "c" denotes it is the third strongest, and "d" denotes it is the fourth strongest. If two or more correlations are denoted by the same letter, this indicates that there is no statistically significant difference in the strength between these correlations

*$p < .05$; **$p < .01$; ***$p < .001$

Table 7 Relative associations between over-reporting versus psychopathology and M-FAST scores

	M-FAST reduced item set total		M-FAST TTL reduced item set total	
Variable	β	p	β	p
Fp-r	0.13	0.053	− 0.03	0.752
EID	0.18	0.004	− 0.01	0.923
THD	0.55	< 0.001	0.56	< 0.001
BXD	− 0.12	0.016	− 0.12	0.102
PTSD sev	0.10	0.115	0.12	0.170

Overall model R^2 for the M-FAST reduced item set total score = .61 ($p < .001$) and for the TTL model = .33 ($p < .001$)

M-FAST Miller Forensic Assessment of Symptoms Test, *TTL* testing-the-limits, *Fp-r* Infrequent Psychopathology, *EID* Emotional/Internalizing Disorders, *THD* Thought Disorders, *BXD* Behavioral/Externalizing Disorders, *PTSD* posttraumatic stress disorder, *sev* severity

negatively associated with M-FAST) but were not significantly related to the testing-the-limits total score. PTSD symptom severity was not significantly associated with either M-FAST score when the other scales were included in the model.

Discussion

Since the time that PTSD was formally recognized in the third edition of the *Diagnostic and Statistical Manual* (*DSM-III*, American Psychiatric Association [APA], 1980), critics have suggested that some individuals may attempt to malinger or feign trauma exposure and PTSD symptoms for secondary gain (Arbisi et al., 2006; Frueh et al., 2005, 1996; Knoll & Resnick, 2006; Rogers et al., 1992). This may be because PTSD is the only major mental health diagnosis in which psychiatric symptoms are causally linked to an environmental antecedent (i.e., trauma exposure) and because individuals may be compensated for trauma exposure, psychopathological responses to trauma, and subsequent disability. PTSD may also be a mitigating factor in criminal cases or grounds for damages in civil ones. Symptom assessments of PTSD tend to be fairly face valid (Elhai, Ford, & Naifeh, 2010), and the PTSD diagnostic criteria are readily available to the general public, making it possible for individuals motivated to feign symptoms of the diagnosis to attempt to do so. These concerns have led to considerable debate about the veracity of reported PTSD symptoms, particularly in veteran populations, where service-related PTSD diagnoses may result in disability benefits from the Federal government (Frueh et al., 1996, 2005, Frueh, Grubaugh, Elhai, & Buckley, 2007; Guriel & Fremouw, 2003; Jackson et al., 2011; Marx et al., 2008, 2012; McNally & Frueh, 2012). Frueh et al. (2005) found that they could not corroborate patient reports of Vietnam combat exposure for nearly 60% of a veteran sample (total $n = 100$) seen in a VA PTSD clinic and raised concerns about compensation seeking in the population. However, larger scale investigations based on military record reviews have failed to find evidence consistent with falsification of trauma exposure or service among veteran samples (Dohrenwend et al., 2006; VA Office of Inspector General, 2005). Given this debate, reliable and valid assessment of both PTSD and malingering among veterans is of critical importance.

The M-FAST is a widely accepted and utilized index of symptom over-reporting and its use carries substantial implications for individuals and broader patient populations alike. Results of this study raise concerns about its use in veteran PTSD samples because the measure showed questionable reliability, included several poorly performing items, and was unduly influenced by psychopathology, even after other markers of symptom over-reporting were accounted for in the analysis. Though the measure was developed as a screening tool, the nature of these concerns raise doubt about its efficiency for screening purposes because (false) positive scores on the measure will lead to unnecessary and time-consuming follow-up evaluations. This concern is consistent with the findings of a recent meta-analysis that concluded that independent investigations of the M-FAST did not yield the same degree of support for the measure's ability to identify symptom exaggeration when compared to data from the original validation samples, leading the authors to caution against stand-alone use of the M-FAST total score (Detullio, Messer, Kennedy, & Millen, 2019). Moreover, both empirical and clinical decision-making efforts to improve the reliability and validity of the measure failed in this study. Eliminating poorly performing items did not substantially alter the pattern of associations with external measures of psychopathology and use of a clinical testing-the-limits procedure yielded a scale that was less strongly influenced by psychopathology, but failed to show improved associations with external measures of symptom over-reporting. The testing-the-limits procedure also yielded scores that were no more reliable than the original scores. These concerns are particularly worrisome given that malingering is defined in the *DSM-5* as "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" (APA,

2013, pg. 726), but these results suggest that over-reporting of symptoms may be a function of unintended misunderstanding of items or idiosyncratic interpretations of them. This sentiment is echoed by Young (2017), who suggested that malingering prevalence may be over-estimated and that the term should be reserved only for instances in which the evidence for this behavior is overwhelming.

Our analysis provides information about why the M-FAST did not reliably and validly assess over-reporting in the veteran PTSD population. We found that, on average, 8.4% of item scores across subjects were changed in the follow-up interview and that simply asking those participants why they had endorsed a particular item led them to discontinue item endorsement (on average, 60% of changed item scores) because they failed to correctly hear or understand the item when it was initially presented. This observation is consistent with evidence of a correlation between MMPI-2-RF TRIN and M-FAST scores. This association could be reflective of difficulty understanding the M-FAST items or a perceived subject expectation to endorse items affirmatively among veterans. However, it is important to note that elevated TRIN has also been shown to be associated with an oppositional response set (e.g., consistent with symptom exaggeration; Anestis, Finn, Gottfried, Arbisi, & Joiner, 2015). In our study, there were also instances in which the participant continued to endorse the item, but further questioning revealed that the symptom was entirely plausible, that the individual responded to only part of the item, or was otherwise unclear about the meaning of the item. Consistent with this, when examining the relative associations between MMPI-2-RF validity and psychiatric symptoms with M-FAST scores, we found that odd and unusual cognitive symptoms (as reflected in the MMPI-2-RF Higher-Order THD scale), and not PTSD symptoms, were the strongest predictor of M-FAST total scores, and THD was far stronger in association than was the best MMPI-2-RF over-reporting validity scale (Fp-r). This suggests that individuals prone to odd and usual cognitive styles and perceptions may be more likely to misunderstand some M-FAST items or otherwise respond to them in an idiosyncratic fashion. In samples of veterans producing valid MMPI-2-RF profiles, RC8 (Aberrant Experiences, which is a contributor to THD) was shown to be related to PTSD reexperiencing symptoms (Wolf et al., 2008) and THD was associated with the dissociative subtype of PTSD in particular (Guetta et al., 2019). This further suggests the relevance of THD to PTSD samples and may help to explain its association with M-FAST scores in this sample.

Some researchers (Guy et al., 2006) have argued that the M-FAST need not be reliable with respect to internal consistency because the measure is multidimensional in nature. However, this argument is problematic because it is commonly understood that reliability is necessary (though not sufficient) for validity. Furthermore, even a multidimensional assessment of an overarching construct requires sufficient homogeneity to provide evidence that all scales or subscales map onto that overarching construct (Clark & Watson, 2019). Rather, it appears that some MFAST items are simply less precise in their ability to capture the malingering construct and that the measure has insufficient true-score variance to be used to reliably assess malingering. This raises questions about the value of published estimates of sensitivity and specificity for veteran PTSD samples. As noted by Lilienfeld, Thames, and Watts (2013), estimates of criterion validity derived from simulation designs may reflect the maximal accuracy of the measure under ideal conditions and therefore may have limited generalizability to applied samples. Specifically, Lilienfeld et al. suggested that extending conclusions from simulation designs to applied samples is based on flawed logic because the finding that simulators produce higher scores on validity scales does not necessarily mean that all high scores are an indication of malingering.

Results of this study suggest the need to identify alternative approaches to assess malingering in PTSD samples. One promising new measure is The Inventory of Problems-29 (IOP-29; Viglione, Giromini, & Landis, 2017), which is a brief self-report scale designed to provide coverage of multiple types of malingering strategies and malingering of both cognitive and psychiatric symptoms; the measure also assesses responses to the test itself, which may be indicative of a problematic response set. The instrument has gone through multiple validation studies in large samples of simulator (including those feigning PTSD), known group, and clinical (including PTSD) samples; classical test construction techniques were used to refine the measure. It has been tested using different types of malingering strategies and random patterns of responding (Giromini, Viglione, Pignolo, & Zennaro, 2019; Viglione et al., 2017) and was recently found to accurately identify simulators from depressed patients, particularly when used in combination with the MMPI-2-RF (Giromini et al., 2019). A unique feature of the scoring is the use of a "false disorder probability score" as opposed to a total or standardized score. Using this score, the sensitivity and specificity of the measure have been shown to be approximately .80 across many different samples and study designs (Viglione et al., 2017; Giromini, Viglione, et al., 2019). One major potential benefit of a probability score is that it assumes an underlying dimensional construct, rather than the categorical one that is inherent in the thresholding approach. The former approach is better aligned with the notion of a spectrum of malingering, ranging from unintentional to intentional behavior (Lilienfeld et al., 2013). Though the measure has been evaluated in PTSD samples (Viglione et al., 2017), its performance in a sample of veterans with a high prevalence of PTSD, specifically, is unknown and requires further investigation.

An alternative approach may be to ask individuals about the extent to which they think their data or

responses to clinical interviews will be used for other purposes (e.g., for disability or legal claims), as we did with the PE Scale. Though beyond the focus of this study, this approach may yield useful information concerning a participant's assumptions about the use of their data, which may necessitate additional follow-up and clarification. Finally, because not all M-FAST items performed poorly, it may be useful to return to well-performing items and generate additional items in the same vein that might expand the coverage and reliability of the measure. For example, several of the items in the Rare Combinations subscale (e.g., items 16–18, 22), as well as two of the Unusual Hallucinations items (20 and 21) and the Unusual Symptom Course item (14) performed quite well in the IRT analyses as evidenced by peaked IICs and steep ICCs. Items 16 and 18 were not significantly higher in those with PTSD, unlike most of the items in the measure, suggesting their discrimination from psychopathology. Furthermore, less than 10% of participants changed their scores on each of these items during the testing-the-limits procedure, further suggesting their robustness. The common feature across these items may be the unusual nature of the symptoms they capture. Regardless, the best assessments of malingering will be multi-model in nature and include self-report, record review, interview, and informant information for evaluating both malingering and PTSD (e.g., Ali et al., 2015; Young, 2017).

Limitations

These results should be interpreted in light of a number of study limitations. The study was based entirely on a veteran sample comprised primarily of men and results may not generalize to other PTSD (or other clinical) samples. We were also unable to examine potential demographic differences in the psychometric properties of the M-FAST as a function of race, ethnicity, or sex, because we did not have sufficient sample size in each sub-group to evaluate this issue. The M-FAST was developed as a screening tool in forensic samples and it is possible that its performance remains acceptable in that population even though it does not perform well in a PTSD sample. Our measure of PTSD was based on an interview-administration of a self-report measure and not the gold standard structured diagnostic interview. Similarly, we did not administer the SIRS-2, the most well-established malingering interview, due to time constraints. Inclusion of the SIRS-2, would have allowed us to compare the incremental utility of the M-FAST, SIMS, and MMPI-2-RF validity scales for the prediction of SIRS-2-defined malingering. Comparisons against the SIMS were limited by poor internal consistency in some subscales (much like the M-FAST). Because the study was designed to evaluate the M-FAST, we did not perform testing-the-limit interviews following positive responses on the MMPI-2-RF validity scales or the SIMS; thus, it is unknown the extent to which some of the concerns raised regarding the M-FAST would generalize to these other measures. Further, the percentage of veterans identified as over-reporting on the SIMS in this study was remarkably high and raises serious doubt about the performance of that measure in this population; a more thorough examination of this issue is necessary in future studies. Because subjects received a small payment for study participation, we could not eliminate all potential financial incentives to malinger, though participants were informed at the outset (both orally and in the written informed consent form) that results of the study would not be included in patient medical records, nor would they be made available to patients or their providers. Finally, we performed a large number of different types of statistical tests without a replication sample, highlighting the need to replicate these findings before unequivocally recommending cessation of M-FAST administration in veteran PTSD samples.

Conclusions

Although literature reviews and recommendations for the detection of malingering in PTSD samples are abundant (e.g., Hall & Hall, 2007; Knoll & Resnick, 2006), the empirical basis for these recommendations is often less than convincing and suggests the need to proceed with caution. Our results raise doubt about the generalizability of results obtained from simulator research designs to clinical samples. In the latter, participants are generally expected to answer honestly and they may have substantially greater or more complex psychiatric symptom severity and impairment than could be feigned by college students, making it more difficult to discriminate true psychopathology from malingering. This highlights the need for new and preferably multi-modal approaches to identify malingering in PTSD samples. The results of this study suggest that despite its widespread use, the M-FAST may be insufficient for this purpose and that simply eliminating poorly performing items or conducting follow-up interviews may likewise be inadequate for addressing concerns related to the measure's reliability, validity, and utility in veteran PTSD samples.

Funding Information Funding for this work was supported by an award from the University of Minnesota Press, Test Division, and by a Presidential Early Career Award for Scientists and Engineers (PECASE 2013A) administered by the US Department of Veterans Affairs (VA) Office of Research and Development, both to Erika J. Wolf. The contents of this manuscript do not represent the views of the US Department of Veterans Affairs or the United States Government.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Ali, S., Jabeen, S., & Alam, F. (2015). Multimodal approach to identifying malingered posttraumatic stress disorder: A review. *Innovations in Clinical Neuroscience, 12*, 12. Retrieved from https://www.ncbi.nlm.nih.gov/

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders (3rd ed.)*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anestis, J. C., Finn, J. A., Gottfried, E., Arbisi, P. A., & Joiner, T. E. (2015). Reading the road signs: The utility of the MMPI-2 Restructured Form Validity Scales in prediction of premature termination. *Assessment, 22*, 279–288. https://doi.org/10.1177/1073191114541672.

Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. (2006). The ability of the MMPI-2 to detect feigned PTSD within the context of compensation seeking. *Psychological Services, 3*, 249–261. https://doi.org/10.1037/1541-1559.3.4.249.

Arbisi, P. A., Polusny, M. A., Erbes, C. R., Thuras, P. R., & Reddy, M. (2011). The Minnesota Multiphasic Personality Inventory–2 Restructured Form in National Guard soldiers screening positive for posttraumatic stress disorder and mild traumatic brain injury. *Psychological Assessment, 23*, 203–214. https://doi.org/10.1037/a0021339.

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form): Manual for Administration and Scoring*. Minneapolis, MN: University of Minnesota Press.

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress, 28*, 489–498. https://doi.org/10.1002/jts.22059.

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders–fifth edition (PCL-5) in veterans. *Psychological Assessment, 28*, 1379–1391. https://doi.org/10.1037/pas0000254.

Brady, R. E., Bujarski, S. J., Feldner, M. T., & Pyne, J. M. (2017). Examining the effects of alexithymia on the relation between posttraumatic stress disorder and over-reporting. *Psychological Trauma: Theory, Research, Practice, and Policy, 9*, 80. https://doi.org/10.1037/tra0000155.

Buckley, T. C., Galovski, T., Blanchard, E. B., & Hickling, E. J. (2003). Is the emotional Stroop paradigm sensitive to malingering? A between-groups study with professional actors and actual trauma survivors. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies, 16*(1), 59–66. https://doi.org/10.1023/A:1022063412056.

Butcher, J. N., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI-2*. Minneapolis: University of Minnesota Press.

Carone, D. A. (2009). Test review of the medical symptom validity test. *Applied Neuropsychology, 16*, 309–311. https://doi.org/10.1080/09084280903297883.

Clark, A. L., Amick, M. M., Fortier, C., Milberg, W. P., & McGlinchey, R. E. (2014). Poor performance validity predicts clinical characteristics and cognitive test performance of OEF/OIF/OND veterans in a research setting. *The Clinical Neuropsychologist, 28*, 802–825. https://doi.org/10.1080/13854046.2014.904928.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319. https://doi.org/10.1037/1040-3590.7.3.309.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*. https://doi.org/10.1037/pas0000626.

Constans, J. I., Kimbrell, T. A., Nanney, J. T., Marx, B. P., Jegley, S., & Pyne, J. M. (2014). Over-reporting bias and the modified Stroop effect in Operation Enduring and Iraqi Freedom veterans with and without PTSD. *Journal of Abnormal Psychology, 123*, 81–90. https://doi.org/10.1037/a0035100.

Constans, J. I., McCloskey, M. S., Vasterling, J. J., Brailey, K., & Mathews, A. (2004). Suppression of attentional bias in PTSD. *Journal of Abnormal Psychology, 113*, 315–323. https://doi.org/10.1037/0021-843X.113.2.315.

Detullio, D., Messer, S. C., Kennedy, T. D., & Millen, D. H. (2019). A meta-analysis of the Miller Forensic Assessment of Symptoms Test (M-FAST). *Psychological Assessment, 31*, 1319–1328. https://doi.org/10.1037/pas0000753.

Dohrenwend, B. P., Turner, J. B., Turse, N. A., Adams, B. G., Koenen, K. C., & Marshall, R. (2006). The psychological risks of Vietnam for US veterans: A revisit with new data and methods. *Science, 313*, 979–982. https://doi.org/10.1126/science.1128944.

Elhai, J. D., Ford, J. D., & Naifeh, J. A. (2010). Assessing trauma exposure and posttraumatic morbidity. In G. M. Rosen & B. C. Frueh (Eds.), *Clinician's guide to posttraumatic stress disorder* (pp. 119–151). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118269961.ch6.

Frueh, B., Grubaugh, A., Elhai, J., & Buckley, T. C. (2007). US Department of Veterans Affairs disability policies for posttraumatic stress disorder: Administrative trends and implications for treatment, rehabilitation, and research. *American Journal of Public Health, 97*, 2143–2145. https://doi.org/10.2105/AJPH.2007.115436.

Frueh, B. C., Elhai, J. D., Grubaugh, A. L., Monnier, J., Kashdan, T. B., Sauvageot, J. A., et al. (2005). Documented combat exposure of US veterans seeking treatment for combat-related post-traumatic stress disorder. *The British Journal of Psychiatry, 186*, 467–472. https://doi.org/10.1192/bjp.186.6.467.

Frueh, B. C., Smith, D. W., & Barker, S. E. (1996). Compensation seeking status and psychometric assessment of combat veterans seeking treatment for PTSD. *Journal of Traumatic Stress, 9*, 427–439. https://doi.org/10.1002/jts.2490090303.

Giromini, L., Lettieri, S. C., Zizolfi, S., Zizolfi, D., Viglione, D. J., Brusadelli, E., Perfetti, B., Carlo, D. A., & Zennaro, A. (2019). Beyond rare-symptoms endorsement: A clinical comparison simulation study using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) with the Inventory of Problems-29 (IOP-29). *Psychological Injury and Law, 12*, 1–13. https://doi.org/10.1007/s12207-019-09357-7.

Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2019). An Inventory of Problems–29 sensitivity study investigating feigning of four different symptom presentations via malingering experimental paradigm. *Journal of Personality Assessment*, 1–10. https://doi.org/10.1080/00223891.2019.1566914.

Glassmire, D. M., Tarescavage, A. M., & Gottfried, E. D. (2016). Likelihood of obtaining Structured Interview of Reported Symptoms (SIRS) and SIRS-2 elevations among forensic psychiatric inpatients with screening elevations on the Miller Forensic Assessment of Symptoms Test. *Psychological Assessment, 28*, 1586. https://doi.org/10.1037/pas0000289.

Goodwin, B. E., Sellbom, M., & Arbisi, P. A. (2013). Posttraumatic stress disorder in veterans: The utility of the MMPI–2–RF Validity Scales in detecting overreported symptoms. *Psychological Assessment, 25*, 671–678. https://doi.org/10.1037/a0032214.

Green, P. (2004). *Manual for the medical symptom validity test*. Green's Publishing Inc..

Guetta, R. E., Wilcox, E. S., Stoop, T. B., Maniates, H., Ryabchenko, K. A., Miller, M. W., & Wolf, E. J. (2019). Psychometric properties of the dissociative subtype of PTSD scale: Replication and extension in

a clinical sample of trauma-exposed veterans. *Behavior Therapy, 50*, 952–966. https://doi.org/10.1016/j.beth.2019.02.003.

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

Guriel, J., Yañez, T., Fremouw, W., Shreve-Neiger, A., Ware, L., Filcheck, H., & Farr, C. (2004). Impact of coaching on malingered posttraumatic stress symptoms on the M-FAST and the TSI. *Journal of Forensic Psychology Practice, 4*, 37–56. https://doi.org/10.1300/J158v04n02_02.

Guriel-Tennant, J., & Fremouw, W. (2006). Impact of trauma history and coaching on malingering of posttraumatic stress disorder using the PAI, TSI, and M-FAST. *The Journal of Forensic Psychiatry & Psychology, 17*, 577–592. https://doi.org/10.1080/14789940600895838.

Guy, L. S., Kwartner, P. P., & Miller, H. A. (2006). Investigating the M-FAST: Psychometric properties and utility to detect diagnostic specific malingering. *Behavioral Sciences & the Law, 24*, 687–702. https://doi.org/10.1002/bsl.706.

Guy, L. S., & Miller, H. A. (2004). Screening for malingered psychopathology in a correctional setting: Utility of the Miller-Forensic Assessment of Symptoms Test (M-FAST). *Criminal Justice and Behavior, 31*, 695–716. https://doi.org/10.1177/0093854804268754.

Hall, R. C., & Hall, R. C. (2007). Detection of malingered PTSD: An overview of clinical, psychometric, and physiological assessment: Where do we stand? *Journal of Forensic Sciences, 52*, 717–725. https://doi.org/10.1111/j.1556-4029.2007.00434.x.

Jackson, J. C., Sinnott, P. L., Marx, B. P., Murdoch, M., Sayer, N. A., Alvarez, J. M., & Owen, R. R. (2011). Variation in practices and attitudes of clinicians assessing PTSD-related disability among veterans. *Journal of Traumatic Stress, 24*, 609–613. https://doi.org/10.1002/jts.20688.

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

Kenny, D. A. (1987). *Statistics for the social and behavioral sciences*. Boston, MA: Little, Brown and Company.

Kilpatrick, D. G., Resnick, H. S., Baber, B., Guille, C., & Gros, K. (2011). *The National Stressful Events web survey (NSES-W)*. Charleston, SC: Medical University of South Carolina.

Knoll, J., & Resnick, P. J. (2006). The detection of malingered posttraumatic stress disorder. *Psychiatric Clinics of North America, 29*, 629–647. https://doi.org/10.1016/j.psc.2006.04.001.

Lewis, J. L., Simcox, A. M., & Berry, D. T. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment, 14*, 170. https://doi.org/10.1037/1040-3590.14.2.170.

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. New York, NY: Oxford University Press.

Lilienfeld, S. O., Thames, A. D., & Watts, A. L. (2013). Symptom validity testing: Unresolved questions, future directions. *Journal of Experimental Psychopathology, 4*, 78–87. https://doi.org/10.5127/jep.028312.

Marion, B. E., Sellbom, M., & Bagby, R. M. (2011). The detection of feigned psychiatric disorders using the MMPI-2-RF over-reporting validity scales: An analog investigation. *Psychological Injury and Law, 4*, 1–12. https://doi.org/10.1007/s12207-011-9097-0.

Marx, B. P., Miller, M. W., Sloan, D. M., Litz, B. T., Kaloupek, D. G., & Keane, T. M. (2008). Military-related PTSD, current disability policies, and malingering. *American Journal of Public Health, 98*, 773–775. https://doi.org/10.2105/AJPH.2007.133223.

Marx, B.P., Jackson, J.C., Schnurr, P.P., Murdoch, M., Sayer, N.A., Keane, T.M., Friedman, M.J., Greevy, R.A., Owen, R.R., Sinnott, P.L. and Speroff, T. (2012). The reality of malingered PTSD among veterans: Reply to McNally and Frueh (2012). *Journal of Traumatic Stress, 25*, 457–460. https://doi.org/10.1002/jts.21714.

McNally, R. J., & Frueh, B. C. (2012). Why we should worry about malingering in the VA system: Comment on Jackson et al. (2011). *Journal of Traumatic Stress, 25*, 454–456. https://doi.org/10.1002/jts.21713.

Merten, T., Lorenz, R., & Schlatow, S. (2010). Posttraumatic stress disorder can easily be faked, but faking can be detected in most cases. *German Journal of Psychiatry, 13*, 140–149. Retrieved from https://psycnet.apa.org/

Miller, H. A. (2001). *Miller forensic assessment of symptoms test: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Miller, H. A. (2004). Examining the use of the M-FAST with criminal defendants incompetent to stand trial. *International Journal of Offender Therapy and Comparative Criminology, 48*, 52–280. https://doi.org/10.1177/0306624X03259167.

Miller, H. A. (2005). The Miller-Forensic Assessment of Symptoms Test (M-FAST) test generalizability and utility across race literacy, and clinical opinion. *Criminal Justice and Behavior, 32*, 591–611. https://doi.org/10.1177/0093854805278805.

Miller, M. W., Wolf, E. J., Kilpatrick, D., Resnick, H., Marx, B. P., Holowka, D. W., & Friedman, M. J. (2013). The prevalence and latent structure of proposed DSM-5 posttraumatic stress disorder symptoms in US national and veteran samples. *Psychological Trauma: Theory, Research, Practice, and Policy, 5*, 501. https://doi.org/10.1037/a0029730.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Office of Inspector General. (2005). Review of state variances in VA disability compensation payments (Report# 05–00765-137).

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *SIRS structured interview of reported symptoms: A professional manual*. Odessa, FL: Psychological Assessment Resources.

Rogers, R., Robinson, E. V., & Gillard, N. D. (2014). The SIMS screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences & the Law, 32*(4), 455–466. https://doi.org/10.1002/bsl.2131.

Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured interview of reported symptoms professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Sellbom, M., & Bagby, R. M. (2010). Detection of overreported psychopathology with the MMPI-2 RF form validity scales. *Psychological Assessment, 22*, 757–767. https://doi.org/10.1037/a0020825.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law Online, 25*, 183-189. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9213290.

Tellegen, A., & Ben-Porath, Y. S. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form): Technical manual*. Minneapolis, MN: University of Minnesota Press.

Veazey, C. H., Wagner, A. L., Hays, J. R., & Miller, H. A. (2005). Validity of the Miller Forensic Assessment of Symptoms Test in psychiatric inpatients. *Psychological Reports, 96*, 771–774. https://doi.org/10.2466/pr0.96.3.771-774.

Viglione, D. J., Giromini, L., & Landis, P. (2017). The development of the Inventory of Problems–29: A brief self-administered measure for discriminating bona fide from feigned psychiatric and cognitive complaints. *Journal of Personality Assessment, 99*, 534–544. https://doi.org/10.1080/00223891.2016.1233882.

Vitacco, M. J., Jackson, R. L., Rogers, R., Neumann, C. S., Miller, H. A., & Gabel, J. (2008). Detection strategies for malingering with the Miller Forensic Assessment of Symptoms Test: A confirmatory

factor analysis of its underlying dimensions. *Assessment, 15*, 97–103. https://doi.org/10.1177/1073191107308085.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2014). The PTSD checklist for DSM-5 (PCL-5). *Scale available from the National Center for PTSD at www.ptsd.va.gov*.

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

Widows, M. R., & Smith, G. P. (2005). *SIMS: Structured Inventory of Malingered Symptomatology*. Lutz, FL: Psychological Assessment Resources, Inc..

Wisdom, N. M., Callahan, J. L., & Shaw, T. G. (2010). Diagnostic utility of the structured inventory of malingered symptomatology to detect malingering in a forensic sample. *Archives of Clinical Neuropsychology, 25*, 118–125. https://doi.org/10.1093/arclin/acp110.

Wolf, E. J., Miller, M. W., Orazem, R. J., Weierich, M. R., Castillo, D. T., Milford, J., & Keane, T. M. (2008). The MMPI-2 Restructured Clinical Scales in the assessment of posttraumatic stress disorder and comorbid disorders. *Psychological Assessment, 20*, 327–340. https://doi.org/10.1037/a0012948.

Young, G. (2017). PTSD in court III: Malingering, assessment, and the law. *International Journal of Law and Psychiatry, 52*, 81–102. https://doi.org/10.1016/j.ijlp.2017.03.001.