A Psychometric Evaluation of the Revised Clinician-Administered PTSD Scale for *DSM-5* (CAPS-5-R) in a Trauma-Exposed Community Sample

by

Brianna Nicole Jackson

A thesis submitted to the Graduate Faculty of
Auburn University
in fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 5, 2023

Keywords: trauma, psychometric evaluation, assessment

Approved by
Chris Correia, Chair, Professor, Department of Psychological Sciences
Tracy Witte, Professor, Department of Psychological Sciences
Frank Weathers, Professor Emeritus, Department of Psychological Sciences

Abstract

The Clinician-Administered PTSD Scale for *DSM-5* (CAPS-5) is a widely used, well-validated structured diagnostic interview for PTSD. The CAPS-5 was recently revised to address limitations of the current instrument. This thesis presents a review of current psychometric data on the CAPS-5 and the results of a psychometric evaluation of the revised instrument, known as the revised Clinician-Administered PTSD Scale for *DSM-5* (CAPS-5-R). Participants were 73 community members with mixed-trauma exposure (e.g., sexual/physical assault, transportation accident, unexpected death of a loved one). CAPS-5-R diagnosis revealed strong interrater reliability (κ = .86 to .93, depending on the scoring rule), test-retest reliability (κ = .73 to .79, depending on the scoring rule), and alternate forms reliability with the CAPS-5 (κ = .79 to .93, depending on the scoring rule). CAPS-5-R total severity scores revealed high interrater reliability (ICC = .98), test-retest reliability (ICC = .86), and alternate forms reliability with the CAPS-5 (ICC = .94). Further, the CAPS-5-R demonstrated strong convergent validity with other measures of PTSD (e.g., PCL-5, DAPS-2) and good discriminant validity with relevant external correlates (e.g., depression, anxiety, alcohol problems, somatic concerns, mania). The results of this study provide promising evidence that the CAPS-5-R is a psychometrically sound measure of *DSM-5-TR PTSD* and a valuable update of the current CAPS-5.

Table of Contents

A Psychometric Evaluation of the Revised Clinician-Administered PTSD Scale for *DSM-5* (CAPS-5-R) in a Trauma-Exposed Community Sample

The Clinician-Administered PTSD Scale for *DSM-5* (CAPS-5; Weathers et al., 2013) is a widely used, well-validated structured diagnostic interview for PTSD. Developed in 1989, the CAPS has since undergone several revisions to maintain concordance with the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) updates and remains widely recognized as the "gold standard" criterion measure of PTSD, against which other new measures are validated. The CAPS-5 has been used in hundreds of research studies to assess PTSD in a wide range of trauma-exposed populations. The CAPS-5 also has been translated into more than a dozen languages and used internationally. In the decade since its release, the CAPS-5 has elicited extensive informal qualitative comments from interviewers regarding its utility, ease of administration, and scoring rules. Based on this feedback, the CAPS-5 was recently revised to streamline administration, clarify prompts and rating guidelines, and create a more granular scoring system. With the recent completion of the revised CAPS-5 (CAPS-5-R), the aim of the present study was to conduct an initial psychometric evaluation in a trauma-exposed community sample. The following sections provide context for this study by reviewing the existing CAPS-5 psychometric literature and describing the rationale for the specific revisions for the CAPS-5-R.

Overview of the CAPS-5

Based on the *DSM-III-R* PTSD criteria, the original CAPS was designed as a comprehensive structured interview to improve upon existing interview measures of PTSD. In developing the CAPS, Blake and coauthors (1995) sought to create an interview that would (a) achieve correspondence with the current *DSM* criteria for PTSD, (b) generate dichotomous and continuous scores of PTSD symptoms, (c) be able to be used by trained paraprofessionals, and

(d) demonstrate strong reliability and validity. As outlined by Weathers and Keane (1999), the CAPS contained several features that were designed to address the limitations of other PTSD interviews. First, in addition to the core symptoms of PTSD, the CAPS also assessed associated symptoms, response validity, overall symptom severity, and the impact of symptoms on social and occupational functioning. Second, the CAPS assessed both the frequency and intensity of each symptom on separate 5-point rating scales, generating continuous and dichotomous scores for each symptom. Third, the CAPS contained behaviorally anchored prompts and rating scales to help increase the reliability of symptom inquiry and severity ratings. Fourth, the CAPS provided specific guidelines for assessing lifetime PTSD diagnostic status, ensuring that the respondent experienced symptoms as part of a syndrome lasting at least one month. Finally, several rationally and empirically derived scoring rules were developed for converting continuous frequency and intensity scores into dichotomous scores. For diagnostic determination, the "F1/I2" rule was developed, whereby items are counted as symptoms toward a diagnosis if the frequency score is 1 or higher and the intensity score is 2 or higher. In addition, frequency and intensity scores were summed to generate item severity scores. These goals and resulting features have continued to provide a foundation for subsequent versions of the CAPS as it has evolved to reflect revisions to the *DSM* PTSD criteria.

Following the publication of *DSM-IV* in 1994, the CAPS authors introduced several substantive changes to the measure, including (a) a protocol for assessing Criterion A (i.e., the stressor criterion, requiring that the index event involves life threat, serious injury, or sexual violence, and engendered fear, helplessness, or horror); (b) explicit prompts and a rating scale to assess trauma-relatedness of nonspecific distress (e.g., sleep difficulties, concentration); (c) replacement of specific associated features with items intended to assess for acute stress

disorder; and (d) general modifications to the intensity scale descriptors. Additional scoring rules were developed based on research conducted by Weathers and colleagues (1999).

The current version of the CAPS, referred to as the CAPS-5, was revised in 2013 to accurately reflect *DSM-5* PTSD criteria. The CAPS-5 introduced several major modifications to the instrument while also retaining the distinctive features of its predecessors. In keeping with *DSM-5* updates, the CAPS-5 (a) eliminated Criterion A2 (i.e., the requirement that the individual experienced fear, helplessness, and horror during the stressor); (b) added three new symptoms (i.e., negative emotional state, self- and other-blame, and reckless behavior); and (c) split avoidance and numbing symptoms into separate clusters. One major update involved implementation of a new scoring system, whereby item frequency and intensity scores are combined into a single 0-4 severity rating, following pre-determined thresholds. This was due primarily to the concern that frequency and intensity ratings were given equal weight, which meant that a high item frequency score could result in a high item severity score even when intensity was low. The new scoring system also included the "SEV2" rule, a slightly more stringent rule for determining whether an item meets the *DSM-5* symptom criterion. Now, for an item to be considered a symptom and counted toward a PTSD diagnosis, it must either have a frequency rating of twice a month or higher and an intensity rating of *Clearly Present* or higher, or a frequency rating of once a month or higher and an intensity rating of *Pronounced* or higher.

In sum, building on previous versions of the CAPS, the CAPS-5 has several notable and distinctive features, including (a) an assessment of Criterion A status, considering life threat, serious injury, and sexual violence; (b) item prompts which assess behavioral referents; (c) a quantitative scoring system based on frequency and intensity of symptoms; (d) explicit rules for combining frequency and intensity to generate item-level severity scores; (e) global ratings of

distress, functional impairment, symptom improvement, and response validity; (f) supplemental items for assessing dissociative reactions; and (g) additional prompts for assessing trauma-relatedness of symptoms that are not inherently linked to the index trauma (e.g. sleep difficulties, loss of interest; Weathers et al., 2018).

Review of the CAPS-5 Psychometric Literature

The CAPS-5 is widely used in the field of traumatic stress and there is extensive evidence of its reliability and validity. This section summarizes the relevant literature to document that the CAPS-5 is a psychometrically sound measure of PTSD diagnostic status and symptom severity.

Literature Review and Study Selection

An initial list of studies was generated using PTSDpubs, an online database produced by the National Center of PTSD, by using the phrase "Clinician-Administered PTSD Scale for *DSM-5*" in the "Tests & Measures" index of the advanced search function. Book chapters, dissertations/theses, conference abstracts, and letters to the editor were excluded. This initial search yielded a total of 264 citations. Further exclusion criteria included duplicate studies and studies in which CAPS-5 psychometric data were not included, which narrowed the list to 96 studies included in the initial review. Studies were subsequently sorted into one of three categories: (a) dedicated psychometric studies in which the CAPS-5 was the primary instrument investigated, (b) non-dedicated psychometric studies in which at least one psychometric index of the CAPS-5 was indicated, and (c) treatment outcome studies in which the CAPS-5 was the primary outcome measure.

Dedicated Psychometric Studies. In 2018, Weathers and colleagues published an initial psychometric evaluation of the CAPS-5 in two independent samples of military veterans (*n*s = 165 and 207). Internal consistency was high for the full scale (α=.88) and somewhat lower and

more variable for the four symptom clusters (α=.55-.77). Interrater reliability was similarly high for PTSD diagnosis (κ =.78), as well as total severity score (ICC=.91), although it should be noted that interrater reliability was slightly higher for CAPS-IV total severity score (ICC=.97). The authors proposed that the lower value for CAPS-5 may have been due to the restricted range of CAPS-5 scores (0 to 80) relative to CAPS-IV scores (0 to 136). Regarding test-retest reliability, Weathers et al. (2018) utilized different interviewers at each administration, thus modeling two sources of error and generating more rigorous indices of reliability more precisely known as coefficients of stability and equivalence. Test-retest reliability was high for PTSD diagnosis (κ =.83) and adequate for total severity score (ICC=.78) and the four symptom clusters (ICC=.64-.80). Convergent and discriminant validity of the CAPS-5 was demonstrated through stronger associations with other measures of PTSD (i.e., PTSD Checklist for *DSM*-5 [PCL-5], PTSD Checklist – Civilian Version [PCL-C]) and weaker associations with measures of other constructs (i.e., anxiety, depression, somatization, disability, and functional impairment). Regarding structural validity, the seven-factor Hybrid model of PTSD was revealed to best fit the data.

Several translated versions of the CAPS-5 have been validated in various diverse populations. Boeschoten et al. (2018) evaluated the psychometric properties of a Dutch-translated version of the CAPS-5 in a community sample of 669 participants. Cronbach's alpha coefficients indicated high internal consistency for the total severity score (α=.90) and adequate levels of internal consistency for the four symptom clusters (α=.63-.79). Interrater reliability was very high for total severity score (ICC=.98), but moderate for diagnostic agreement (Fleiss' κ=.59). To evaluate structural validity of the CAPS-5, Boeschoten and colleagues (2018) utilized

CFA to compare six models of *DSM-5* PTSD, concluding that the six-factor Anhedonia model was the best fit to the data.

Boysan et al. (2017) validated a Turkish version of the CAPS-5 in a sample of 90 participants. Internal consistency of the symptom clusters was high (α=.87-.94). Symptom endorsement was not significantly different between two separate administrations, demonstrating test-retest reliability. Concurrent validity was demonstrated through significant differences in PTSD symptom endorsement by cluster among three groups (i.e., PTSD-positive, depressive-only, healthy controls), wherein individuals with PTSD were more likely to endorse at least one symptom in each cluster. Further, CAPS-5 symptom clusters demonstrated robust associations with measures of related constructs (PCL-5, Dissociative Experiences Scale [DES], Beck Depression Inventory [BDI], Beck Anxiety Inventory [BAI]), providing strong evidence for the convergent validity of the CAPS-5. An evaluation of the structural validity of the CAPS-5 revealed that, although the *DSM-5* four-factor model of PTSD provided adequate fit to the data, the six-factor Externalizing Behaviors model of PTSD fit the data best.

Lastly, Müller-Engelmann and colleagues (2020) validated a German version of the CAPS-5 in two independent samples: women with PTSD after childhood physical/sexual abuse (*n*=174) and community members (*n*=100). Results demonstrated high internal consistency for total severity score (α=.93), but lower and more variable coefficients for the subscale scores (α=.65-.85). Interrater reliability of the total severity score was also high (ICC=.81 for seven raters with respect to two administrations; ICC=.89 for a second rater with respect to twelve administrations). The CAPS-5 was highly correlated with a different measure of PTSD, the Posttraumatic Diagnostic Scale (PDS; r=.87), and less correlated with the BDI-II (r=.72),

demonstrating convergent and discriminant validity. Regarding structural validity, the seven-factor Hybrid model of PTSD demonstrated the best fit to the data.

Non-Dedicated Psychometric Studies. The CAPS-5 has been the primary measure of PTSD in numerous research studies. For inclusion in this section of the review, studies had to report at least one index of reliability (i.e., internal consistency, interrater reliability, or test-retest reliability) or validity (i.e., convergent and discriminant validity, discriminative validity, or structural validity). Forty-six studies were determined to meet these criteria; however, due to space constraints, only the most representative studies are discussed in detail below.

In a validation study of the Cognitive Emotional Regulation Questionnaire (CERQ), Lee et al. (2020) employed the CAPS-5 in community clinical (n=244) and male veteran (n=236) samples. Internal consistency was found to be high for the CAPS-5 total severity score (α=.85). Interrater reliability for CAPS-5 PTSD diagnosis was similarly high (κ =.80). Regarding validity, a series of Pearson correlations was conducted to evaluate associations between the six subscales of the CERQ and CAPS-5 total severity score. Significant positive associations were found with Self-Blame (r = .36) and Perseveration (r = .26). Lower, non-significant associations were revealed for Acceptance (r = .04), Other-Blame (r = .02), and Reappraisal (r = -.13). Finally, a significant negative association was revealed for Positive Refocusing (r = -.15).

The CAPS-5 was also utilized in a separate validation study evaluating the psychometric properties of the Posttraumatic Avoidance Behavior Questionnaire (PABQ) in a sample of 177 treatment-seeking adults (Shor et al., 2021). Interrater reliability for PTSD diagnosis was found to be high (κ =.85). A series of Pearson correlations were utilized to evaluate associations between the subscales and total scores of the PABQ and CAPS-5. Effect sizes were generally moderate to large, with the avoidance and NACM clusters of the CAPS-5 demonstrating the

largest associations with the PABQ total score. However, the PABQ subscale intended to assess avoidance of trauma-related thoughts evidenced the weakest associations with CAPS-5 total and subscale scores.

In a large prospective study of 490 Black female hospital patients, Maples-Keller et al. (2021) examined the association between personality traits and PTSD symptoms. CAPS-5 total and cluster scores were correlated with the Personality Inventory for *DSM-5* (PID-5) scales. The largest associations were found between CAPS-5 scores and the Negative Affectivity domain of the PID-5 (r = .40-.58), followed by the Psychoticism (r = .39-.50), Detachment (r = .37-.49), Disinhibition (r = .27-.41), and Antagonism (r = .17-.25) domains. Maples-Keller et al. (2021) also performed a multiple regression analysis and found that only Negative Affectivity (β=.30), Psychoticism (β=.14), and Detachment (β=.11) were significant predictors of the CAPS-5 total score.

Using a subset (n=190) of Maples-Keller et al.'s (2021) sample, Powers et al. (2017) investigated differential predictors of *DSM-5* PTSD and ICD-11 Complex PTSD. Participants were sorted into three groups: *DSM-5* PTSD (assessed via the CAPS-5; n = 35), ICD-11 Complex PTSD (assessed via the ICD-11 Trauma Interview; n = 33), and Traumatized Control (n = 118). Regarding interrater reliability, 12 CAPS-5 interviews were selected for review and found to have high diagnostic agreement (κ =.83). When comparing groups on measures of relevant constructs, individuals in the *DSM-5* PTSD group had significantly higher scores on measures of depression, emotion dysregulation, and dissociation than individuals in the Traumatized Control group.

In a sample of 125 treatment-seeking veterans, Martin et al. (2021) utilized the CAPS-5 to investigate the role of PTSD symptoms in the association between thwarted belongingness,

perceived burdensomeness, and suicidal ideation. Regarding reliability, internal consistency of the CAPS-5 full scale was found to be high (α=.84). A series of Pearson correlations were conducted to examine associations between CAPS-5 total severity score and measures of relevant constructs. As evidence of convergent and discriminant validity, CAPS-5 scores correlated r = .46 with the Patient Health Questionnaire-9 (PHQ-9), r = .41 with the perceived burdensomeness scale on the Interpersonal Needs Questionnaire (INQ-15), r = .27 with the Beck Scale for Suicidal Ideation (BSS), r = .26 with the thwarted belongingness scale of the INQ-15, and r = .001 with the substance use disorder scale of the Structured Clinical Interview for *DSM-5* (SCID-5).

As part of a larger study examining the association between negative trauma-related cognitions and pain-related functioning, Curry et al. (2019) administered the CAPS-5 to 137 veterans with comorbid PTSD and Alcohol Use Disorder (AUD). Participants were also administered measures of pain-related disability and pain severity (assessed via the Pain Disability Questionnaire [PDQ]), negative trauma-related cognitions (assessed via the Posttraumatic Cognitions Inventory [PTCI]), alcohol use frequency, and AUD severity (assessed via the SCID-IV). Regarding reliability, Cronbach's alpha of the CAPS-5 total score was found to be high (α=.83). As evidence of validity, CAPS-5 total score correlated r =.58 with the PTCI total score, r = .42 with AUD severity, r = .30 with pain-related disability, r = .10 with pain severity, and r = .03 with alcohol use frequency.

Foa et al. (2016) used the CAPS-5 in their evaluation of the psychometric properties of the Posttraumatic Stress Disorder Symptom Scale Interview for *DSM-5* (PSSI-5). As evidence of convergent validity, the CAPS-5 total score revealed large associations with the PSSI-5 total score (r = .72), the PDS-5 (r = .64), and the PTSD Checklist-Stressor Specific (PCL-S; r = .63).

Weaker associations were observed between the CAPS-5 total score and the State-Trait Anxiety Inventory (r = .56) and the BDI-II (r = .57), providing evidence for the discriminant validity of the CAPS-5. Diagnostic agreement between the CAPS-5 and PSSI-5 was lower (κ=.49), potentially due to the small sample size or to the diagnostic unreliability of the PSSI-5 (Foa et al., 2016; Weathers et al., 2018).

Treatment Outcome Studies. The CAPS-5 has been utilized in numerous treatment studies as the primary outcome measure of PTSD. Forty-four studies were deemed initially eligible for inclusion in this section. Six studies were excluded due to overlapping samples and failing to provide pre- and post-treatment data, resulting in a total of 38 studies. Studies were categorized by whether the primary treatment evaluated was (a) psychosocial, (b) pharmacological, or (c) neither psychosocial nor pharmacological (i.e., transcranial stimulation, electroconvulsive therapy, respiratory guided intervention).

As outlined in Weathers et al. (2001), several important points should be considered when assessing the CAPS-5's sensitivity to clinical change. First, virtually any treatment outcome study will find a reduction in CAPS-5 scores from pre- to post-treatment. This is due to several factors, including placebo effects, sample requirements (i.e., participants selected on the basis of extreme scores), and repeated assessment as an intervention in and of itself. Therefore, studies with comparison groups are especially important in evaluating the CAPS-5's sensitivity to change because we should expect to see greater improvement in more potent treatment conditions. Second, studies varied significantly based on how CAPS-5 outcome data was reported. Following Weathers et al.'s (2001) example, percent change in CAPS-5 total score was calculated utilizing pre- and post-treatment scores when provided, enhancing comprehensibility and allowing for comparisons across studies. Further, studies also varied by the number of

measurement points included (i.e., pre-treatment, mid-treatment, post-treatment, follow-up), therefore only percent change in PTSD from pre- to post-treatment (i.e., immediately following treatment) was presented (see Table 1 for a summary of key findings).

Within-groups effects. Regarding psychosocial interventions, there was a significant reduction in total CAPS-5 scores in all 10 studies (Studies 3, 7–9, 14–16, 18, 20, and 22). Improvement in total CAPS-5 scores ranged from 17-73%, with a median of 49%. Regarding pharmacological interventions (regardless of whether it was supplemented with psychotherapy), only one study did not utilize a comparison group (Study 31). In this study, there was a significant reduction of 69% in CAPS-5 total score at post-treatment; however, once depressive symptoms were controlled for, significance was eliminated. Lastly, regarding other treatment studies, there was a significant reduction of CAPS-5 total scores in a study using theta-burst stimulation (Study 34) and in a study using electroconvulsive therapy (Study 38), ranging from 22% to 36%.

Between-groups effects. Of the 27 psychosocial intervention studies, 14 were randomized controlled trials (RCTs; Studies 1, 2, 4, 6, 11, 13, 17, 19, 21, 23, 24–27), one was a matched pairs design with partial randomization (Study 10), one was a crossover design (Study 5), and one was a quasi-experimental design (Study 12). Of the RCTs, six studies found significant between-groups effects (Studies 1, 2, 6, 11, 19, 26) in favor of the more potent psychosocial intervention. Six studies did not find a significant between-groups effect (Studies 4, 17, 23, 24, 25, 27), although it should be noted that two of these studies (Studies 25 and 27) aimed to test the noninferiority of a novel intervention when compared to Cognitive Processing Therapy (CPT), a gold-standard PTSD intervention. Study 13 found a significant between-groups effect for the portion of their sample with comorbid PTSD and MDD, but not for

participants without an MDD diagnosis. Lastly, Study 21 did not examine between-groups differences at post-treatment (rather, only at the one-year follow-up), but revealed comparable improvements across all conditions at post-treatment. Of the remaining studies, Study 10, which utilized a matched pairs design, found no significant between-groups effects; Study 12, which utilized a quasi-experimental design, found a significant between-groups effect for the more active intervention. Lastly, Study 5, which utilized a crossover design, did not conduct an analysis of between-groups differences, but improvements were comparable between both interventions.

Of the pharmacological interventions, four studies were randomized placebo-controlled trials (Studies 28–30, and 33), and one study was an RCT comparing the efficacy of two drugs (Study 32). Two of the placebo-controlled studies found significant between-groups effects, with greater improvement observed for the drug conditions (Studies 28 and 29), while one did not (Study 33). Study 32 also found a significant between-groups effect for the Ketamine versus Midazolam condition. Lastly, Study 30 did not find a significant between-groups effect for any of the drug conditions compared to placebo.

Of the remaining interventions utilizing other treatment methods, two studies were randomized sham-controlled trials (Studies 35 and 36), and one study was an RCT comparing the efficacy of two interventions (one trauma specific and the other non-trauma specific; Study 37). Of the sham-controlled trials, one study did not find a significant between-groups effect (Study 35) and the other study found a significant between-groups effect in favor of the placebo condition (Study 36). Finally, Study 37 also did not find a significant between-groups effect.

Discussion

The preceding review indicates that the CAPS-5 is a psychometrically sound measure for generating continuous and dichotomous ratings of PTSD in clinical and research settings. Interrater reliability for CAPS-5 continuous scores ranged from .81 to .98, indicating high to very high consistency. Interrater reliability for CAPS-5 PTSD diagnosis was generally around a kappa of .80, indicating substantial agreement among raters. The one exception was Boeschoten et al.'s (2018) validation of the Dutch-translated CAPS-5, which revealed a considerably lower kappa value of .59. Boeschoten et al. (2018) proposed several explanations for the lower interrater reliability, including what they characterized as a reduction in scoring specificity from the CAPS to CAPS-5, presumably allowing more opportunities for scoring inconsistencies. Only two studies (Weathers et al., 2018 and Boysan et al., 2017) evaluated test-retest reliability of the CAPS-5, revealing strong evidence for consistency between testing occasions. Regarding internal consistency, alphas were consistently in the .80 to .90 range for CAPS-5 total scores and in the .60 to .90 range for the four symptom clusters. The fact that the Avoidance cluster of the CAPS-5 is restricted to only two items may explain the generally lower and more variable alphas of this symptom cluster.

While more difficult to integrate and evaluate, a review of the literature supports the construct validity of CAPS-5 by demonstrating conceptually consistent associations with several relevant constructs. Regarding convergent and discriminant validity, CAPS-5 correlations with other measures of PTSD (i.e., PCL-5, PDS-5, PSSI-5) were generally high, ranging from .52 to .87. More importantly, associations with measures of other related constructs (i.e., depression, anxiety, generalized distress) were comparatively lower, and associations with measures of conceptually unrelated constructs (i.e., alcohol use, pain severity, psychopathy) were non-significant. Taken together, findings indicate that the CAPS-5 is able to effectively discriminate

between PTSD and conceptually related constructs such as depression and anxiety. Studies examining group differences also lend support to the discriminative validity of CAPS-5 dichotomous ratings of PTSD. Individuals determined to meet criteria for PTSD via the CAPS-5 revealed significantly higher scores on measures of generalized distress, depression, anxiety, emotion dysregulation, and dissociation than those who did not meet criteria for PTSD.

Results from the review of treatment outcome studies provide evidence for the CAPS-5's ability to detect symptom change within groups; however, findings are mixed regarding between-groups differences. As discussed earlier, there are many factors that may contribute to significant improvement in symptoms after an intervention, including placebo effects, sample characteristics, and repeat assessments, which may inflate findings. Indeed, all but one uncontrolled trial revealed a significant reduction in CAPS-5 symptoms at post-treatment. In contrast, only about half of the controlled psychosocial and pharmacological trials revealed significant between-groups effects. It is unclear whether these mixed findings are due to the inefficaciousness of certain interventions or due to limitations regarding the CAPS-5's sensitivity to clinical change.

Limitations of the CAPS-5

As it stands, the CAPS-5 is a reliable and valid measure for generating multiple indices of posttraumatic stress severity. However, it has some limitations related to its scoring and administration that make it challenging to learn and may reduce its psychometric performance. These limitations are described in detail below.

Scoring

As outlined previously, one significant update from the CAPS-IV to the CAPS-5 involved a change in the scoring method. In previous versions of the CAPS, severity scores were

calculated by adding the frequency and intensity scores of individual items, both of which are rated on a 5-point (0-4) scale, to create a single score ranging from 0-8 at the item level, and 0-136 at the total severity level. The CAPS-5 departed from this scoring method due to insufficient conceptual justification for giving equal weight to frequency and intensity when identifying clinically meaningful symptoms (Weathers et al., 2018). Instead, a rationally and empirically derived scoring method based on previous research (Weathers, Ruscio, & Keane, 1999) was implemented to convert frequency and intensity scores into a single 5-point (0-4) severity rating, resulting in a total PTSD severity score ranging from 0-80. An item severity score of 2 is counted as a symptom. This requires either (a) a minimum frequency rating of *twice a month* and a minimum intensity rating of *clearly present*, or (b) a minimum frequency rating of *once a month* and a minimum intensity rating of *pronounced*.

From a conceptual standpoint, the current CAPS-5 scoring method provides a better indication of clinically relevant symptoms than its predecessor. However, an unanticipated tradeoff involves the now truncated range in item-level severity scores. From a statistical analysis perspective, this reduction in scale granularity sacrifices data precision, as many statistical procedures rely heavily on the assumption of variability of data (Pearse, 2011). This has implications for CAPS-5 item-level analyses, which are further compounded by a preponderance of severity ratings of 2=*Moderate* and 3=*Severe*. Moreover, in the event that raters are between two ratings, they are instructed to give the lower rating. Yet despite this explicit rule, raters may be inclined to choose a rating based on their clinical judgment, even if this results in deviation from the lowest-rating rule, which ultimately reduces scale reliability and validity. An alternative solution is needed to increase scale precision while also maintaining the conceptual rigor of the current CAPS-5 scoring method.

Explication of PTSD Symptoms

The primary goal of the CAPS-5 is to assess *DSM-5* PTSD criteria with fidelity; however, issues regarding operationalization of symptoms are apparent. This poses a problem for raters with only a superficial knowledge of trauma and PTSD, who may rate items based on their limited understanding of the PTSD criteria. For particular items, the lack of clarity regarding intended meaning of symptoms may impact the reliability of ratings.

Item B3 (*Dissociative reactions [e.g., flashbacks] in which the individual feels or acts as if the traumatic event[s] were recurring. [Such reactions may occur on a continuum, with the most extreme expression being a complete loss of awareness of present surroundings])* is an example of such an item. While the given prompts aim to help the rater discriminate between strong memories and flashbacks, assessment of dissociative qualities (i.e., sights, smells, sounds) is limited. Given that many respondents may report transient awareness of surroundings and behavioral changes during powerful memories, and not necessarily flashbacks, it may be difficult to distinguish between the two occurrences.

Within the Negative Alterations in Cognition and Mood (NACM) cluster, there are several items that are particularly susceptible to scoring incongruities. The first (and most controversial regarding endorsement rate and reliability) is item D1 (*Inability to remember an important aspect of the traumatic event[s] [typically due to dissociative amnesia, and not to other factors such as head injury, alcohol, or drugs]*). The initial prompt of this item asks the respondent to indicate whether they have had difficulty remembering important parts of the index event. Several optional, follow-up prompts are given to discriminate between dissociative amnesia and other factors, such as a head injury, substance use, or normal forgetting. Respondents may positively endorse this item until they are asked to consider these other factors,

thereby prolonging the interview and increasing respondent burden. Additionally, respondents are asked to provide a discrete number of "important aspects" of the event that they have had difficulty remembering. Providing an answer to this prompt can be frustrating and difficult for respondents, who have likely never needed to conceptualize their posttraumatic amnesia in this way. In addition, the frequency rating for this item introduces a certain level of ambiguity for the rater. For example, what is considered an "important aspect?" If two respondents answer that they have difficulty recalling one important aspect, yet one respondent indicates that their index event contained only two important aspects whereas the other respondent indicates that their index event contained many important aspects, should their frequency rating be the same? Responses to this prompt may be difficult to comprehend and score.

Item D2 (*Persistent and exaggerated negative beliefs or expectations about oneself, others, or the world [e.g., "I am bad," "No one can be trusted," "The world is completely dangerous," "My whole nervous system is permanently ruined"]*) presents with similar issues. Unlike other items which intend to assess for negative trauma-related feelings and behaviors, item D2 aims to evaluate the presence of negative core beliefs related to the index event. For raters with limited knowledge of core beliefs, it may be difficult to distinguish between feelings, behaviors, and beliefs. While optional follow-up prompts are offered within parentheses, they have the potential to be skipped despite providing potentially useful information. Further, intensity ratings for this item are based on both the degree of distortion and conviction of the belief; however, there is little indication of this in the provided scoring guidelines.

Item D3 (*Persistent, distorted cognitions about the cause or consequences of the traumatic event[s] that lead the individual to blame himself/herself or others*) is especially susceptible to scoring discrepancies for reasons similar to those outlined in D2. Given that many

of the follow-up prompts are within parentheses, and thus considered optional, raters may fail to administer them despite their potential usefulness.

Lastly, item E3 (*Hypervigilance*) would benefit from additional prompts intended to capture the full range hypervigilant behaviors in different situations. While the initial prompt asks the respondent to provide examples of watchful behaviors, the lack of specific probing regarding safety rituals and watchful behaviors in different locations precludes a well-informed rating. This is especially needed, as scoring guidelines denote location (i.e., in public versus at home) and exaggerated behaviors (i.e., rituals, routines) as important indicators of intensity. Without this information, raters may need to rely on clinical judgment when assigning a symptom rating.

A common issue among all CAPS-5 items is the large number of follow-up prompts that are italicized, offset, and in parentheses and thus may be considered by raters to be optional. While the intention was likely to streamline administration, raters may be quick to forego administering these prompts even when they may benefit from the additional information. For several items, such as D1, the parenthetical prompts may solicit necessary information related to the presence of the symptom but instead risk the chance of being disregarded.

Respondent Burden

The average length of administration for the CAPS-5 is 45 to 60 minutes (Weathers et al., 2018), varying considerably based on the number of symptoms endorsed, the verbosity of the respondent, and the proficiency of the rater. Coupled with the extensive length of administration is the amount of effort required of the respondent to provide satisfactory answers (i.e., responses that are able to inform an accurate rating). While fixed factors such as tangentiality of responses and limited cognitive resources of the respondent can negatively impact the validity of the test

results, there is one feature of the CAPS-5 that can be targeted to reduce respondent burden and increase accuracy of given responses.

The process of obtaining respondent estimates of symptom frequency can have implications for respondent burden. As outlined in Figure 1, the CAPS-5 has four thresholds for determining a frequency rating (with the exception of item D1): less than twice a month or less than 20% (subthreshold), twice a month or greater than 20% (threshold), twice a week or greater than 50% (markedly elevated), and daily/almost daily or greater than 80% (incapacitating). For a greater proportion of items, respondents are asked to rate "how often" they have been bothered by the symptom in the past month. In the absence of response anchors, respondents may provide answers that are either insufficient (e.g., 20 "times") or difficult to score (e.g., daily for one week, and then once a week for the remaining three weeks). Regarding the percentage estimates of frequency (i.e., "how much of the time in the past month have you felt that way, as a percentage?"), respondents may misunderstand the intended meaning (e.g., "I feel [that way] very strongly: 100%") or have difficulty conceptualizing their experience using percentages. As a result, raters may be compelled to restate the question or provide response anchors (e.g., "would you say more than twice a month, but less than twice a week?"), at best introducing anchoring bias, and at worst supplying a response. Throughout this process, the respondent may begin to feel frustrated, exhausted, and self-conscious, which may reduce the validity of subsequent responses.

Rater Burden

Although all limitations of an instrument will have implications for rater burden, there are specific areas of the CAPS-5 that can be targeted to mitigate the impact of rater burden on obtaining valid and reliable estimates of PTSD.

While the CAPS-5 includes a scoring grid to provide more specific guidance for combining intensity and frequency ratings into the appropriate severity score (see Figure 2), the scoring grid was developed several years after the publication of the CAPS-5 and is therefore not included in the original interview document. The scoring grid is now an included component of the CAPS-5; however, users who adopted the CAPS-5 prior to the development of the scoring grid may not be aware of its existence. This is unfortunate, as the scoring grid may reduce the time and cognitive energy expended for calculating severity ratings and enhance the accuracy of PTSD severity ratings by reducing the risk of scoring errors.

Alongside the previously discussed limitations related to explication of PTSD symptoms, the formatting of item prompts may be especially confusing when distinguishing between required, recommended, and optional prompts. With just a cursory look at the interview, one can observe an array of bolded, un-bolded, parenthetical, and italicized typeface. Consequently, item prompts risk being skipped or administered unnecessarily.

Explanation of Revisions to the CAPS-5

Specific improvements for the CAPS-5-R fall into five different areas: (a) simplified formatting, (b) revised prompts, (c) expanded rating options, (d) addition of a frequency rating card, and (e) scoring guidelines and grid. What follows is an enumeration of these revisions to the CAPS-5, henceforth referred to as the revised Clinician-Administered PTSD Scale for *DSM-5*, or CAPS-5-R.

Simplified Formatting

The formatting of the CAPS-5 was simplified to improve visual flow, allowing for easier navigation through the sequence of prompts, and clarity regarding which prompts are required and which are optional. Many prompts which were previously italicized, offset, and/or

parenthetical, and thus may have been considered optional by raters, are now required to maximize information gathered.

Revised Prompts

Item prompts of the CAPS-5 were revised to enhance comprehensibility and ensure that all necessary information is gathered to make a reliable rating. This included reworking several existing prompts and adding new prompts to create a more detailed inquiry for areas that had proven to be ambiguous in the CAPS-5. For example, prompts assessing the degree of symptom-related distress were reworded to reduce vague responses and elicit more specific behavioral examples from the respondent. Extensive changes were also made regarding the conceptualization of the more challenging items, such as those outlined in the limitations section.

Expanded Rating Options

Intensity and frequency rating options were expanded to allow for more precise estimates of symptom severity. The CAPS 5's item-level scoring method involves large intervals between frequency and intensity ratings (see Table 1 for current frequency and intensity ratings), which may reduce the impact of frequency ratings on overall symptom severity (for example, an item with a frequency of seven times a month and an intensity rating of *Pronounced* would be assigned a severity score of 2 despite being closer to a rating of 3). A similar effect can be seen when assigning intensity ratings which are between ratings of *Clearly Present, Pronounced,* and *Extreme.* Given that raters are instructed to assign the lower rating, it may be possible for a respondent to have a "true" symptom severity closer to a rating of 3, yet receive a 2 due to not meeting the established cut-offs for behavioral referents. To mitigate these issues and increase reliability of symptom severity scores, intermediate intensity ratings were added between *Clearly*

Present and *Pronounced,* and between *Pronounced* and *Extreme*, and frequency thresholds were expanded.

Previously measured on a 0–4 scale, symptom severity scores are now measured on a 0–10 scale. On its face, this solution seems to have the effect of overcomplicating an already complex scoring system; however, the benefits of adopting this revision, especially for research considerations, are clear. More precise data and increased variability in symptom severity scores may produce more meaningful results when statistically analyzed (Pearse, 2011). Additionally, a natural consequence of increased scale granularity is the ability to more closely detect symptom-level change over a period of time (Lim et al., 2016). Rather than overhauling the CAPS-5 scoring method, previously established cut-off rules were implemented for ease of translating scores back to a 0–4 scale, allowing for backward compatibility between the CAPS-5 and CAPS-5-R (see Figure 3). By increasing scale precision, the burden of relying on clinical judgment to decide between two ratings may be diminished, thereby increasing interrater reliability for continuous scores.

In the absence of specific research on CAPS-5 scoring limitations, an evaluation of early psychometric studies provides some support for the hypothesis that increasing scale width will increase reliability of a measure. Based on Symonds' (1924) proposition, early researchers theorized that a scale with too few anchors fails to discriminate between responses, thus introducing random errors that lead to underestimates of reliability indices (e.g., internal consistency, interrater, test-retest; Weng, 2004). Indeed, research suggests that increasing scale width results in a steady increase in interrater reliability up to seven anchors, at which point reliability coefficients plateau (Cicchetti et al., 1985; Symonds, 1924). Other early researchers have found evidence that providing more scale anchors increases other reliability indices, such as

test-retest reliability and internal consistency (Cook & Beckman, 2009; Oaster, 1989; Weng, 2004). A recent study using IRT analysis suggested a higher number of response categories yield better estimates of reliability (Daher et al., 2015). Furthermore, Korte et al. (2013) assessed whether revisions to item scoring enhanced the psychometric properties of a measure of nicotine dependence. The authors concluded that increased scale anchors resulted in significant improvement in scale reliability.

A significant reason for utilizing dimensional rating scales is to attempt to quantify categorical constructs, the success of which depends on ensuring adequate representation among responses (Pearse, 2011). Increasing the precision of the scales may not only increase the statistical reliability of the CAPS-5-R, but also ensure that the CAPS-5-R can sufficiently capture and discriminate between the precise range of PTSD reactions.

Frequency Rating Card

In an effort to reduce respondent burden, a frequency rating card was created in order to efficiently assess symptom frequency. This card is given to the respondent at the beginning of the interview and referred to by the respondent when determining the frequency of their symptoms, thus eliminating the risk of introducing clinical judgment to an otherwise objective aspect of symptom severity. The frequency rating card may also enhance respondent engagement and reduce the duration of interview.

Scoring Guidelines and Grid

Revisions to the CAPS-5 also include the addition of scoring guidelines and a scoring grid on every page. Regarding scoring guidelines, each item now includes more detailed guidance for determining intensity ratings. It is expected that these guidelines facilitate more reliable intensity ratings by reducing reliance on clinical judgment, particularly in instances

where the rater is not as familiar with the conceptual underpinnings of PTSD criteria. A scoring grid for determining symptom severity is now included at the bottom of every page, reducing the likelihood of scoring mistakes and facilitating a more efficient rating process.

Aims

As discussed in the previous section, the CAPS-5-R represents a significant improvement over the current version, with the potential to increase the reliability and validity of PTSD ratings. However, the psychometric properties of the CAPS-5-R have yet to be empirically investigated. Accordingly, the current investigation aimed to assess the psychometric properties of the CAPS-5-R. Specifically, this study assessed the CAPS-5-R's internal consistency, test-retest and interrater reliability, convergent and discriminant validity, backward compatibility with the CAPS-5, and rater receptiveness.

Aim 1: Provide descriptive statistics (i.e., range, mean, standard deviation, frequency) of scores generated using the CAPS-5-R.

Aim 2: Evaluate the internal consistency, test-retest reliability, and interrater reliability of the CAPS-5-R.

Aim 3: Evaluate the convergent and discriminant validity of the CAPS-5-R's dichotomous (present/absent) and dimensional item, cluster, and total scale scores through associations with the CAPS-5 and other relevant external correlates.

Aim 4: Determine clinician receptiveness to the CAPS-5-R, given the substantial changes in scoring and item prompts.

Method

Participants and Procedure

Psychometric study

The study sample consisted of 76 adult community members who were recruited from the southeastern region of the United States as part of a larger validation protocol. Participants were recruited from flyers posted around the community and social media advertisements. The Auburn University institutional review board (IRB) reviewed and approved the following procedures. All participants completed an online screener survey to determine initial eligibility for study inclusion. Participants were invited to participate in the study if they endorsed experiencing a *DSM-5-TR* Criterion A traumatic event, determined by their responses on the Life Events Checklist-Extended Version (LEC-5–Extended; Weathers et al., 2013), and at least one symptom of PTSD, assessed via the Primary Care PTSD Screen for *DSM-5* (PC-PTSD-5; Prins et al., 2016). Exclusion criteria included (a) learning about the death of a loved one before the age of 5 years old as the index event; (b) refusing to allow recording of the interview session; (c) for those who preferred to complete the interview remotely, not having access to the internet, a webcam, or a confidential space; and (d) refusing to provide emergency contact information.

Once initial eligibility was determined and informed consent was obtained, participants completed an online battery of self-report questionnaires, followed by two clinical interview sessions scheduled approximately one week apart (*M* days = 8.7, *SD* = 3.1). All interviews were administered by advanced clinical psychology doctoral students. Participants were randomly assigned to either a test-retest condition (CAPS-5-R administered twice) or alternate forms condition (CAPS-5 and CAPS-5-R administered in counterbalanced order across participants). They were compensated a total of $100 upon completion of all study procedures. Three participants were excluded from the final analyses based on invalid Personality Assessment

Inventory (PAI) responses, using Morey's (1996) cut-offs (INF ≥ 75*T*, ICN ≥ 73*T*, PIM ≥ 68*T*). This resulted in a final sample size of 73.

See Table 2 for sociodemographic characteristics of all participants. The final sample consisted of primarily female (70%) participants with a mean age of 42.6 years (range=19–71, *SD* =13.1). The racial distribution of the sample was majority White (83.6%), followed by Black (11%), American Indian/Alaskan Native (1.4%), and other race (4.1%). The most prevalent traumas endorsed were sexual assault (24.7%), physical assault (21.9%), sudden, unexpected death of a loved one (20.5%), and transportation accident (9.6%). Based on the *DSM-5-TR* PTSD diagnostic rule (≥ 1 reexperiencing symptom, ≥ 1 avoidance symptom, ≥ 2 negative alterations in cognition and mood symptoms, ≥ 2 alterations in arousal and reactivity symptoms), 49% of the sample met criteria for PTSD at Time 1.

Rater feedback

Interviewers (*n*=5) completed anonymous feedback surveys with sets of questions addressing (a) receptiveness to CAPS-5-R changes; (b) perceived improvement of the CAPS-5-R in different domains (e.g., length of administration, ease of scoring); and (c) interview preference (i.e., CAPS-5, CAPS-5-R, or both). Interviewers also completed an open-ended question to explain their ratings in greater detail.

Measures

Primary Measures

Clinician-Administered PTSD Scale for DSM-5 (CAPS-5). The CAPS-5 is a structured clinical interview used to assess PTSD diagnostic status and symptom severity (Weathers et al., 2013). The CAPS-5 contains 20 items that assess core PTSD symptoms based

on *DSM-5* criteria; additional items that measure onset and duration of symptoms, global subjective distress and functional impairment, response validity, and overall severity; and two items that assess depersonalization and derealization for the dissociative subtype of PTSD. Raters assess the intensity and frequency of each symptom, which are then combined to generate an item severity score based on predetermined scoring rules. Intensity is rated as *Minimal, Clearly Present, Pronounced,* and *Extreme.* Frequency is recorded as either the number of times or percentage of time in the recall period (i.e., past month or past week). Item severity scores range from 0 (*Absent*) to 4 (*Extreme/incapacitating*). Items with a severity score of 2 (*Moderate/threshold*) or greater are considered "present" symptoms. A PTSD diagnosis is given when the *DSM-5* criteria are met (i.e., the presence of at least 1 reexperiencing symptom, 1 avoidance symptom, 2 negative alterations in cognition in mood symptoms, and 2 arousal and reactivity symptoms). The CAPS-5 is designed for use in both clinical practice and research. The CAPS-5 has been found to have high internal consistency for the full scale (α=.88) and subscales (α=.55-.77), and high test-retest reliability for diagnostic agreement (κ=.83) and total severity score (ICC=.78; Weathers et al., 2018). In the current sample, internal consistency of the CAPS-5 full scale was high (α = .93).

Clinician-Administered PTSD Scale for DSM-5 Revised (CAPS-5-R). The CAPS-5-R is a structured clinical interview used to assess PTSD diagnostic status and symptom severity (Weathers et al., 2022). The CAPS-5-R is a revised version of the CAPS-5 and has the same structure and number of items as the CAPS-5. However, the intensity rating options were expanded by including intermediate ratings between *Clearly Present* and *Pronounced*, as well as between *Pronounced* and *Extreme*. In addition, frequency was revised so that respondents now utilize a frequency rating card to make frequency estimates. For items in which frequency is

assessed as a "number of times," respondents select one of six options, ranging from *Once a month* to *Daily/almost every day*. For items in which frequency is assessed as a "percentage of the time," respondents select any percentage from 0-100, guided by anchors ranging from *Rarely (0-10%)* to *Nearly all the time (80-100%)*. Following predetermined scoring rules, raters combine intensity and frequency into an item severity score ranging from 0-10. Items rated as 3 or greater are considered "present" symptoms. A PTSD diagnosis is given when certain criteria are met (i.e., the presence of 1 reexperiencing symptom, 1 avoidance symptom, 2 negative alterations in cognition in mood symptoms, and 2 arousal and reactivity symptoms). There are currently no psychometric data on the CAPS-5-R.

Secondary Measures

Mini International Neuropsychiatric Interview (MINI). The MINI is a structured clinical interview for assessing *DSM-5* psychiatric disorders. The MINI is used in both clinical and research settings. Questions are rated on a yes/no scale, wherein interviewees report current and past symptoms of most major psychiatric disorders. Each module begins with screening questions. If the initial criteria are met, a set of follow-up questions are provided to help the interviewer gather more information on the duration and extent of the symptoms. The MINI has excellent interrater reliability ($\kappa > .75$), and very good test-retest reliability, with all kappas above .75 (with the exception of current mania, which was below .40). The MINI has good sensitivity, with a value of .70 or greater for all but three disorders (dysthymia, OCD, and drug dependence), and excellent specificity (greater than .85; Sheehan et al., 1998).

PTSD Checklist for DSM-5 (PCL-5). The PCL-5 is a 20-item *DSM-5*-correspondent questionnaire used to measure PTSD symptom severity (Weathers et al., 2013). Respondents indicate how much they were bothered by each PTSD symptom, using a 5-point Likert scale

ranging from 0 (*not at all*) to 4 (*extremely*). A total symptom severity score is calculated by summing the scores from the individual items, resulting in a composite score range of 0 to 80 with higher scores indicating greater PTSD symptom severity. The PCL-5 has been found to have good test-retest reliability (r = .82-.86), excellent internal consistency for the full scale (α=.94), and good diagnostic utility (Bovin et al., 2016). PCL-5 scores have also demonstrated convergent and discriminant validity with relevant external correlates (Blevins et al., 2015). In the present study, internal consistency of the PCL-5 full scale was high (α=.94).

Multiscale Dissociation Inventory (MDI). The MDI is a 30-item questionnaire designed to assess a wide range of dissociative symptoms (i.e., disengagement, depersonalization, derealization, emotional constriction, memory disturbance, and identity dissociation; Briere, 2002). Frequency of each item is rated on a 5-point Likert scale ranging from 1 (*never*) to 5 (*very often*). MDI scales have been found to have good internal consistency and convergent and discriminant validity with relevant external correlates (Jeffirs, 2023). In the current sample, internal consistency of the MDI was high (α=.96).

The Personality Assessment Inventory (PAI). The PAI is a 344-item multiscale questionnaire used to assess a broad range of psychopathology and personality traits (Morey, 1991). The PAI includes 22 scales designed to assess clinical concerns (i.e., anxiety, depression, anxiety related disorders [e.g., obsessive-compulsive disorder], mania, paranoia, schizophrenia, borderline features, antisocial features, somatic concerns, alcohol problems, and drug problems), treatment consideration (i.e., aggression, suicidal ideation, nonsupport, stress, and treatment rejection), interpersonal functioning (i.e., dominance and warmth), and response validity (i.e., inconsistency, infrequency, positive impression, and negative impression). Items are rated on a 4-point Likert scale (F= *False, not at all true*; ST = *Slightly true*; MT = *Mainly true*; VT = *Very*

true). PAI scale scores have been found to have high internal consistency (αs=.81–.86), excellent test-retest reliability (*r*s =.79–.92) and excellent convergent and discriminant validity (Morey, 2007). The current study only utilized the clinical scales as a measure of convergent and discriminant validity. In the current sample, internal consistency estimates of PAI clinical scales were high (αs=.81–.93).

Detailed Assessment of Posttraumatic Stress-2 (DAPS-2). The DAPS-2 is a 119-item multiscale questionnaire used to assess trauma exposure and related symptomatology (Briere, 2001). The DAPS-2 contains a total PTSD scale, as well as four subscales corresponding to *DSM-5* PTSD criteria. The DAPS-2 also contains additional subscales for assessment of functional impairment, dissociation, suicidality, substance abuse, and response bias. Respondents initially answer a series of yes/no questions regarding whether they experienced a Criterion A traumatic event. Subsequent questions ask participants to indicate whether they have experienced various traumatic stress reactions on a 5-point Likert scale, ranging from 1 (*Not at all*) to 5 (*Very much*). Participants are also instructed to rate the frequency of experiencing these reactions in the last month from 1 (*Never*) to 5 (*4 or more times a week*). DAPS-2 scales have demonstrated very good internal consistency (α=.78-.99) and excellent convergent and discriminant validity (Petri et al., 2020). The current study only utilized the Posttraumatic Stress Total scale (PTS-T) as a measure of convergent validity. The internal consistency of the PTS-T scale was found to be high in the current sample (α=.97).

Data Analysis

To assess the internal consistency of CAPS-5-R cluster and total scores, Cronbach's alpha and McDonald's omega coefficients were calculated and item-total and interitem correlations were examined for all CAPS-5-R Time 1 administrations (*n*=55). Given that alpha

relies on specific assumptions of the instrument (which are rarely met) and is heavily influenced by the number of items in a scale, alpha coefficients may provide an inaccurate and inflated measure of internal consistency. In contrast, McDonald's omega is a more robust estimator of internal consistency given that it assumes a congeneric model (i.e., different factor loadings) and is not as dependent on the number of items in the scale (Hayes & Coutts, 2020). Thus, it is expected that omega coefficients provide a more realistic approximation of internal consistency.

To examine test-retest and interrater reliability of CAPS-5-R dimensional scores, intraclass correlation coefficients (ICC) were calculated at the item, cluster, and scale total level. Regarding test-retest reliability, given that a different rater administered the interview at the retest session, reliability estimates derived from this method are more precisely referred to as coefficients of stability and interrater equivalence because they model two sources of measurement error, viz., testing occasion and rater (Crocker & Algina, 1986). The specific ICC was a one-way random effects model for a single measurement, which estimates absolute agreement between independent measurements (McGraw & Wong, 1996; Shrout & Fleiss, 1979). Based on recommendations from Bujang and Baharum (2017) for calculating intraclass correlation coefficients (ICC), an *a priori* sample size estimation of n=10 is required to detect a moderate effect size of 0.70 when alpha and power are fixed at 0.05 and 0.8, respectively. The final sample size of n=29 participants in both retest conditions provided sufficient power for estimating test-retest and interrater reliability.

To examine test-retest and interrater reliability of CAPS-5-R binary scores (i.e., the presence or absence of symptoms and PTSD diagnosis), Cohen's kappa coefficients were calculated for dichotomous scores at the diagnostic level. Bujang & Baharum (2017) also provide a comprehensive table for calculating the minimum sample size required for estimating

Cohen's kappa. The minimum kappa value expected to be obtained by researchers for diagnostic agreement in the current study was 0.5 ($\kappa 2=.5$) when there is no assumed agreement in the first place ($\kappa 1=0$). The minimum sample size needed to detect a minimum kappa value of 0.5 is $n=29$ when alpha and power are fixed at 0.05 and 0.8, respectively. The final sample size of $n=29$ participants in both retest conditions provided sufficient power for detecting diagnostic agreement between two independent raters and administrations. Following guidelines provided by Landis and Koch (1977), kappas above .60 were interpreted as substantial agreement between raters, and kappas above .80 were interpreted as near-perfect agreement between raters.

To examine the concordance between the CAPS-5-R with the CAPS-5, ICCs were calculated for dimensional scores at the item, cluster, and total scale level, and Cohen's kappa was calculated for dichotomous scores at the diagnostic level. Given that the CAPS-5-R and CAPS-5 reflect different but nearly equivalent forms of the same interview, this concordance was conceptualized as alternate forms reliability. Further, because the CAPS-5-R utilizes a different scale (i.e., 0–10) for assessing dimensional scores, CAPS-5-R scores were recoded into CAPS-5 scores (i.e., 0-4), and ICCs were re-calculated using these recoded CAPS-5-R scores.

To examine the convergent and discriminant validity of the CAPS-5-R with measures of relevant constructs, zero-order correlations were calculated. Following the Westen and Rosenthal (2003) approach for evaluating construct validity, a pattern of correlations was predicted based on existing validity evidence (i.e., using the studies outlined in the literature review section as reference). Effect-size statistics generated from this approach aim to summarize the fit between predicted and observed patterns of correlations. The first statistic, $r_{\text{alerting-CV}}$, is a measure of the level of agreement between the predicted and observed correlations. The second statistic, $r_{\text{contrast-CV}}$, is a more stringent measure of agreement that considers sample size, median intercorrelation

among criterion variables, and the magnitudes of correlations between the measure being validated and criterion variables. Sample sizes were unequal for criterion correlations as a result of missing questionnaire data. Accordingly, the unweighted harmonic mean of *n* was calculated for each set of criterion measures based on recommendations by Rosenthal (see Poythress et al., 2010).

Lastly, to examine informal, qualitative feedback regarding clinician receptiveness to using the CAPS-5-R, rater responses were compiled and scrutinized for emergent themes.

Alpha coefficients, ICCs, kappa coefficients, and zero-order Pearson correlations were calculated using SPSS version 26. Omega coefficients were calculated using *Mplus* Version 7 (Muthén & Muthén, 2017).

Results

Descriptive statistics and internal consistency of all scales and subscales used in the current study are provided in Table 3.

Descriptive Statistics and Internal Consistency

CAPS-5-R descriptive statistics are presented in Table 4. The McDonald's omega (ω = .90) and Cronbach's alpha (α = .89) coefficients were both high for CAPS-5-R full scale. Internal consistency was more variable for the clusters (α = .61 - .77; ω = .61 - .80). In line with previous psychometric findings of the CAPS-5, the internal consistency of the avoidance cluster of the CAPS-5-R was comparably lower than the other three clusters.

Corrected item-total correlations for the full scale ranged from .04 to .71. One symptom, item D1 (amnesia), revealed a very low item-total correlation of .04. Once removed, the item-total correlation of the remaining 19 items ranged from .23 to .71. Corrected item-total correlations were also calculated for each of the four symptom clusters. For the reexperiencing

cluster, corrected item-total correlations ranged from .37 to .63. Both items within the avoidance cluster revealed a corrected item-total correlation of .44. For the NACM cluster, corrected item-total correlations ranged from .15 to .68; although when D1 was removed, corrected item-total correlations ranged from .34 to .68. Lastly, for the alterations in arousal and reactivity cluster, corrected item-total correlations ranged from .20 to .65.

Interitem correlations for the full scale ranged from -.16 to .69, with a mean of .28. Two symptoms, item D1 (amnesia) and E2 (recklessness), revealed very low interitem correlations, ranging from -.16 to .20 for D1 and -.11 to .38 for E2. When these items were removed, the remaining 18 symptoms revealed a mean interitem correlation of .33. Interitem correlations were also calculated for each of the four symptom clusters. Interitem correlations for the reexperiencing cluster ranged from .19 to .51, with a mean of .38. The interitem correlation for the avoidance cluster was .44. For the NACM cluster, interitem correlations ranged from -.02 to .67, with a mean of .32; however, when D1 was dropped, interitem correlations ranged from .13 to .67, with a mean of .40. For the alterations in arousal and reactivity cluster, interitem correlations ranged from .07 to .69, with a mean of .29; however, when E2 was dropped, interitem correlations ranged from .09 to .69, with a mean of .37.

Test-Retest Reliability

A paired samples *t*-test was conducted to examine whether the difference between CAPS-5-R Time 1 total scores (M =53.3, SD = 30.1) and Time 2 total scores (M =55.1, SD = 27.8) significantly differed from zero. Results were not significant, $t(28) = -0.73$, $p = .472$, indicating that scores did not significantly differ between timepoints.

Test-retest reliability was very high for the total severity score (ICC=.86; 95% CI [.73, .93]), and moderate to high for the four symptom clusters (reexperiencing ICC= .81; 95%

CI [.64, .91], avoidance ICC= .66; 95% CI [.40, .83], NACM ICC= .79; 95% CI [.61, .90], and alterations in arousal and reactivity ICC= .82; 95% CI [.65, .91]). Test-retest reliability varied at the item level. Most items revealed moderate to good test-retest reliability (ICC ≥ .50). However, items B4, B5, D4, and E2 revealed below adequate reliability across timepoints (see Table 5).

Test-retest reliability for CAPS-5-R diagnostic agreement was substantial (κ=.73), resulting in 25 of 29 correct classifications. Three participants obtained a diagnosis at Time 1 but not Time 2, and one participant obtained a diagnosis at Time 2 but not Time 1. To calibrate correspondence between the CAPS-5-R the CAPS-5, an additional requirement of a minimum total severity score was added. The highest correspondence between the CAPS-5-R and CAPS-5 was found when a minimum total severity score of 54 for the CAPS-5-R (SEV3/54) and 26 for the CAPS-5 (SEV2/26) was added as a requirement alongside the basic scoring rules. Test-retest reliability for the CAPS-5-R was higher for SEV3/54 (κ=.79), resulting in 26 of 29 correct classifications.

Interrater Reliability

Interrater reliability (assessed via video recordings) was conducted by a third, independent rater who was not involved in the original interview sessions. Interrater reliability was very high for total severity score (ICC=.98; 95% CI [.96, .99]), and for the four symptom clusters (reexperiencing ICC= .96; 95% CI [.92, .98], avoidance ICC= .92; 95% CI [.85, .96], NACM ICC= .97; 95% CI [.93, .98], and alterations in arousal and reactivity ICC= .98; 95% CI [.96, .99]). At the item level, interrater reliability was good to excellent (ICC ≥ .75). However, item scores for assessing derealization revealed only moderate interrater reliability (ICC=.63; see Table 5). Regarding diagnostic agreement, interrater reliability was near-perfect (κ=.86). This

comparison resulted in 27 of 29 correct classifications. Interrater reliability increased when implementing the SEV3/54 scoring rule (κ=.93), resulting in 28 of 29 correct classifications.

Alternate Forms Reliability

When CAPS-5-R scores were recoded into CAPS-5 scoring, the CAPS-5-R revealed excellent concordance with the CAPS-5. At the dimensional level, CAPS-5-R and CAPS-5 scores were strongly associated (total ICC=.94; 95% CI [.88, .97], reexperiencing ICC= .78; 95% CI [.59, .89], avoidance ICC= .86; 95% CI [.72, .93], NACM ICC= .86; 95% CI [.73, .93], and alterations in arousal and reactivity ICC= .82; 95% CI [.65, .91]; see Table 5). Most items revealed moderate to excellent concordance (ICC ≥ .50). However, item E1 revealed below adequate concordance between the CAPS-5-R and the CAPS-5 (see Table 5).

A substantial association on PTSD diagnosis between the CAPS-5-R SEV3 and CAPS-5 SEV2 scoring rules was observed, with a kappa of .79. This resulted in 26 out of 29 correct classifications, with two participants with a diagnosis on the CAPS-5 but not the CAPS-5-R, and one participant with a diagnosis on the CAPS-5-R but not the CAPS-5. An increase in concordance between the CAPS-5-R and the CAPS-5 was observed when using the CAPS-5-R SEV3/54 and CAPS-5 SEV2/26 scoring rules (κ=.93). This resulted in 28 out of 29 correct classifications.

Convergent and Discriminant Validity

The CAPS-5-R correlated as expected with measures of PTSD and constructs theoretically related to PTSD (see Table 6). Regarding convergent validity, the strongest associations were observed between the CAPS-5-R and other measures of PTSD (i.e., PCL-5, DAPS PTS-T, and PAI ARDT). Regarding discriminant validity, the CAPS-5-R correlated

moderately with measures of theoretically related constructs (e.g., dissociation, depression, anxiety, borderline features) and least strongly with measures of unrelated constructs (e.g., alcohol use, drug use, antisocial features). A few discriminant correlations were observed to be higher than expected (i.e., schizophrenia, paranoia, and mania). Construct validity effect size correlations between CAPS-5-R and criterion measures were high (Time 1 interviews only [n=55]: $r_{\text{contrast-CV}} = 0.81$; $r_{\text{alerting-CV}} = 0.88$; alternate forms interviews [n=29]: $r_{\text{contrast-CV}} = 0.98$; $r_{\text{alerting-CV}} = 0.90$).

Rater feedback

Raters expressed mostly positive feedback regarding changes to the CAPS-5-R (see Table 7). The majority of changes were perceived by raters to have at least somewhat improved upon the CAPS-5 in several domains, with the exception of length of administration. Most raters (80%) perceived that the length of administration remains similar between the CAPS-5 and the CAPS-5-R.

Several themes were identified from the free responses. First, raters felt that the addition of the scoring guidelines has eased administration and improved the validity of symptom ratings:

> *"The scoring instructions on the side of the CAPS-5-R have been most helpful when I'm feeling stuck on how to score and have increased my confidence in the rating I provide."*

> *"The more in-depth scoring guidelines at the side of each page have truly been a godsend…it is so easy to forget those simple things (for example, is loss of interest only due to fearful avoidance? Are they reporting specific negative emotions as a result of reminders, or more generally?). I have gotten a ton of use out of those guidelines."*

Raters also felt that the addition of the frequency rating card has greatly reduced both participant and rater burden while also increasing the validity of symptom ratings:

> *"The frequency rating card really eases the collection of frequency and reduces the likelihood of making a scoring error (e.g., collecting a frequency on the CAPS-5 that falls between a severity, such as "1-2x/month"). The frequency rating card also seems to have removed some of the "guessing game" participants may experience when giving frequency, and I've noticed that the response time has been quicker given there are direct anchors."*

> *"With the CAPS-5, it is so easy for raters/participants to get frustrated when trying to get frequency responses, and the rating card just eliminates that entirely. It also eliminates any opportunity to accept "in between" responses (like 1-2x a month)…"*

Regarding the expansion of the rating scale, all raters indicated that the addition of the intermediate ratings has lessened frustration with scoring, particularly in tandem with the scoring grid at the bottom of each page:

> *"The scoring is also so much easier. I find myself far less confused or on the fence with my ratings with the 5-R."*

> *"It was somewhat difficult to transition from the CAPS-5-R back to the CAPS-5 since intensity ratings were more constrained."*

> *"The scoring grid at the bottom of the page is a fantastic edition, reducing additional resources/materials needed on hand to score appropriately (e.g., with the CAPS-5 I have to go and find the separate document of the scoring grid to continue scoring) and reduces length of scoring."*

Lastly, raters indicated that the revisions to the CAPS-5 have greatly decreased participant and rater burden:

> *"…participants have seemed to find the CAPS-5-R less confusing and frustrating in comparison to the CAPS-5."*
>
> *"… the ease of administration, clarity, and length of the measure lessens the burden for both participants and raters using the CAPS-5-R. The updates made, I believe, will aid in the confidence in which the CAPS is used which will undoubtedly reduce burden across the board."*

If given the option to utilize the CAPS-5 or the CAPS-5-R in future clinical and research endeavors, all raters reported that they would be most likely to use the CAPS-5-R.

Discussion

The current study aimed to provide a psychometric evaluation of the CAPS-5-R in a trauma-exposed community sample. Results indicated that the CAPS-5-R is a highly reliable and valid interview measure for generating PTSD severity scores and making PTSD diagnoses. Results also demonstrated strong correspondence between the CAPS-5-R and CAPS-5, with alternate forms reliability for total severity even slightly exceeding test-retest reliability. Finally, results revealed a high degree of rater satisfaction with the CAPS-5-R and a preference for it over the CAPS-5.

CAPS-5-R total and cluster scores demonstrated high internal consistency. Consistent with previous psychometric studies of the CAPS-5, the internal consistency for the avoidance cluster was lower than the other three clusters, likely due to the avoidance cluster being represented by only two items. Mean interitem correlations of the total scale and four clusters

were within the recommended range proposed by Clark and Watson (1995), although item D1 (amnesia) and E2 (recklessness) revealed very low interitem correlations. Item-total correlations of the total scale and four clusters were also found to be adequate, with the exception of item D1. The below-adequate performance of items D1 and E2 may be due to their infrequent endorsement by participants, mirroring similar findings in psychometric evaluations of the CAPS-5.

Test-retest reliability of the CAPS-5-R was substantial at both the diagnostic and dimensional level. Notably, interviews were administered by independent raters at both timepoints, indicating that CAPS-5-R scores are stable across occasions and raters. At the item level, CAPS-5-R scores generally revealed adequate test-retest reliability. However, several items (i.e., B4 [psychological distress], B5 [physiological reactions], D4 [persistent negative emotional state], and E2 [recklessness]) revealed below adequate reliability across timepoints. The relative instability of these items across occasions may be due to several factors, including a focus on more transient experiences that may differ from week to week, discrepancies in participant responses across timepoints, or rater error. However, differences due to rater are likely to be minimal, given that interrater reliability was very high for all of these items.

Interrater reliability of the CAPS-5-R was found to be near-perfect, demonstrating remarkable consistency between raters. The revisions included in the CAPS-5-R, such as increased variability in item scores, clarification of item prompts, and scoring guidelines, likely contributed to substantial agreement between raters. Interrater reliability at the item level was generally excellent, with only one item (derealization) revealing moderate consistency between raters.

Regarding backward compatibility, CAPS-5-R scores demonstrated excellent concordance with CAPS-5 scores at the dimensional level. The diagnostic concordance between the CAPS-5-R and the CAPS-5 using basic scoring rules (i.e., SEV3 for CAPS-5-R and SEV2 with CAPS-5) was found to be substantial. Furthermore, when a minimum total severity score was added as a requirement (54 for CAPS-5-R and 26 for CAPS-5), diagnostic concordance reached a near-perfect level. When comparing CAPS-5-R standard scores to those generated using CAPS-5 scoring, comparable estimates of internal consistency, test-retest and interrater reliability, and convergent and discriminant validity were observed, further supporting the notion of the CAPS-5-R's backward compatibility with the CAPS-5. Results demonstrate that the CAPS-5-R closely corresponds to its predecessor, indicating its suitability as a replacement for the CAPS-5.

Regarding convergent and discriminant validity, the CAPS-5-R correlated most strongly with other measures of PTSD, moderately with measures of related constructs (i.e., dissociation, depression, anxiety, borderline personality features, schizophrenia, paranoia, and somatization), and non-significantly with measures of unrelated constructs (i.e., alcohol problems, drug problems, antisocial personality features). Although there were a few associations that were higher than expected (e.g., schizophrenia, mania, paranoia), the inflated associations with these measures are likely due to shared item content with several symptoms of PTSD (i.e., social detachment, hypervigilance, irritability). Nonetheless, large construct validity effect sizes were found for CAPS-5-R Time 1 interviews. Further, construct validity effect sizes of CAPS-5-R and CAPS-5 within the same sample of participants were highly similar, demonstrating that the CAPS-5-R is a valid measure of posttraumatic stress.

Lastly, this study elicited rater feedback regarding the CAPS-5-R. Raters provided an overwhelmingly positive review of the revised instrument, particularly regarding unique revisions (such as the frequency rating card and scoring guidelines) and their impact on participant and rater burden. One message was resoundingly clear: The CAPS-5-R improves upon its predecessor in nearly every domain, and if given the option, all raters indicated that they would continue to utilize the CAPS-5-R in research and clinical contexts. Given that the revisions to the CAPS-5 were primarily informed by the experiences of raters and clinicians, it will be important for future evaluations of the CAPS-5-R to continue to assess interviewers' perceptions of the instrument.

Results of the present study should be evaluated in light of several limitations. First, sample sizes were modest and may have attenuated findings. However, the purpose of the study was to provide a preliminary psychometric evaluation of the CAPS-5-R using a rigorous and systematic approach. Given the promising results of the current study, it is likely the case that greater sample sizes and increased variability would provide an even more convincing demonstration of the CAPS-5-R's psychometric properties. Second, the majority of individuals in our sample were White. The limited racial/ethnic diversity in our sample precluded further investigation of how these demographic factors may have influenced study results. Lastly, resource limitations precluded a parallel comprehensive psychometric evaluation of the CAPS-5, particularly regarding test-retest reliability. Although incorporating this information would have allowed for more direct comparisons between the CAPS-5 and the CAPS-5-R, it should be noted that the wealth of literature regarding the CAPS-5's psychometric properties establishes a strong standard for evaluating the CAPS-5-R.

In conclusion, this study provides promising evidence that the CAPS-5-R is a psychometrically sound measure of *DSM-5-TR* PTSD and a valuable update of the current CAPS-5. The simplified formatting streamlines administration, the revised prompts facilitate accurate inquiry, the new intermediate Intensity anchors allow more nuanced ratings, and the expanded scoring grid increases the range and variability of item, symptom cluster, and total Severity scores. Further, the CAPS-5-R demonstrated reliability and validity at levels comparable to or even exceeding that of published evidence for the CAPS-5. Finally, the CAPS-5-R demonstrated a high degree of backward compatibility, as evidenced by its strong association with the CAPS-5.

Future directions include replicating these findings in larger samples in other trauma populations and evaluating the CAPS-5-R's sensitivity to clinical change in clinical trials. At this point, researchers and clinicians are encouraged to switch to the CAPS-5-R when feasible. Experienced CAPS raters will appreciate the various improvements and adjust quickly to the new scoring grid, and new CAPS raters will be able to learn standard administration and scoring more efficiently. Further, the backward compatibility permits straightforward integration of new CAPS-5-R data with existing CAPS-5 datasets. This will ensure continuity in the transition to the CAPS-5-R, the latest iteration in the long-standing tradition of the CAPS.

References

Andersen, T. E., Ravn, S. L., Armfield, N., Maujean, A., Requena, S. S., & Sterling, M. (2021). Trauma-focused cognitive behavioural therapy and exercise for chronic whiplash with comorbid posttraumatic stress disorder: A randomised controlled trial. *PAIN*, *162*(4), 1221–1232. https://doi.org/10.1097/j.pain.0000000000002117

Assmann, N., Fassbinder, E., Schaich, A., Lee, C. W., Boterhoven de Haan, K., Rijkeboer, M., & Arntz, A. (2021). Differential effects of comorbid psychiatric disorders on treatment outcome in posttraumatic stress disorder from childhood trauma. *Journal of Clinical Medicine*, *10*(16), 3708. https://doi.org/10.3390/jcm10163708

Bernardi, J., Dahiya, M., & Jobson, L. (2019). Culturally modified cognitive processing therapy for Karen refugees with posttraumatic stress disorder: A pilot study. *Clinical Psychology & Psychotherapy*, *26*(5), 531–539. https://doi.org/10.1002/cpp.2373

Bisson, J. I., van Deursen, R., Hannigan, B., Kitchiner, N., Barawi, K., Jones, K., Pickles, T., Skipper, J., Young, C., Abbott, L. R., van Gelderen, M., Nijdam, M. J., & Vermetten, E. (2020). Randomized controlled trial of multi-modular motion-assisted memory desensitization and reconsolidation (3MDR) for male military veterans with treatment-resistant post-traumatic stress disorder. *Acta Psychiatrica Scandinavica*, *142*(2), 141–151. https://doi.org/10.1111/acps.13200

Blain, R. C., Pukay-Martin, N. D., Martin, C. E., Dutton-Cox, C. E., & Chard, K. M. (2021). Residential cognitive processing therapy decreases suicidality by reducing perceived burdensomeness in veterans with posttraumatic stress disorder. *Journal of Traumatic Stress*, *34*(6), 1199–1208. https://doi.org/10.1002/jts.22618

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, *8*(1), 75–90. https://doi.org/10.1002/jts.2490080106

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*(6), 489-498. https://doi.org/10.1002/jts

Boeschoten, M. A., Van der Aa, N., Bakker, A., Ter Heide, F. J. J., Hoofwijk, M. C., Jongedijk, R. A., Van Minnen, A., Elzinga, B. M., & Olff, M. (2018). Development and evaluation of the Dutch Clinician-Administered PTSD Scale for DSM-5 (CAPS-5). *European Journal of Psychotraumatology*, *9*(1), 1546085. https://doi.org/10.1080/20008198.2018.1546085

Bohus, M., Kleindienst, N., Hahn, C., Müller-Engelmann, M., Ludäscher, P., Steil, R., Fydrich, T., Kuehner, C., Resick, P. A., Stiglmayr, C., Schmahl, C., & Priebe, K. (2020). Dialectical behavior therapy for posttraumatic stress disorder (DBT-PTSD) compared with cognitive processing therapy (CPT) in complex presentations of PTSD in women survivors of childhood abuse: A randomized clinical trial. *JAMA Psychiatry*, *77*(12), 1235–1245. https://doi.org/10.1001/jamapsychiatry.2020.2148

Bonn-Miller, M. O., Sisley, S., Riggs, P., Yazar-Klosinski, B., Wang, J. B., Loflin, M. J. E., Shechet, B., Hennigan, C., Matthews, R., Emerson, A., & Doblin, R. (2021). The short-term impact of 3 smoked cannabis preparations versus placebo on PTSD symptoms: A randomized cross-over clinical trial. *PLOS ONE*, *16*(3), e0246990. https://doi.org/10.1371/journal.pone.0246990

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*(11), 1379–1391. https://doi.org/10.1037/pas0000254

Boysan, M., Guzel Ozdemir, P., Yilmaz, E., Selvi, Y., Özdemir, O., & Celal Kefeli, M. (2017). Psychometric properties of the Turkish version of the Clinician-Administered PTSD Scale for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (Turkish CAPS-5). *Psychiatry and Clinical Psychopharmacology*, *27*(2), 173–184. https://doi.org/10.1080/24750573.2017.1326746

Briere, J., Weathers, F. W., & Runtz, M. (2005). Is dissociation a multidimensional construct? Evidence from the Multiscale Dissociation Inventory. *Journal of Traumatic Stress*, 221–231.

Bryan, C. J., Leifker, F. R., Rozek, D. C., Bryan, A. O., Reynolds, M. L., Oakey, D. N., & Roberge, E. (2018). Examining the effectiveness of an intensive, 2-week treatment program for military personnel and veterans with PTSD: Results of a pilot, open-label, prospective cohort trial. *Journal of Clinical Psychology*, *74*(12), 2070–2081. https://doi.org/10.1002/jclp.22651

Bujang, M. A., & Baharum, N. (2017). A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Journal of School of Dental Sciences*, *12*(1), 1-11.

Bujang, M. A., & Baharum, N. (2017). Guidelines of the minimum sample size requirements for kappa agreement test. *Epidemiology, Biostatistics and Public Health*, *Authors' Manuscripts*. https://doi.org/10.2427/12267

Carletto, S., Porcaro, C., Settanta, C., Vizzari, V., Stanizzo, M. R., Oliva, F., Torta, R., Fernandez, I., Moja, M. C., Pagani, M., & Ostacoli, L. (2019). Neurobiological features and response to eye movement desensitization and reprocessing treatment of posttraumatic stress disorder in patients with breast cancer. *European Journal of Psychotraumatology, 10*(1), 1600832. https://doi.org/10.1080/20008198.2019.1600832

Cicchetti, D. V., Shoinralter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement, 9*(1), 31–36. https://doi.org/10.1177/014662168500900103

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cook, D. A., & Beckman, T. J. (2009). Does scale length matter? A comparison of nine-versus five-point rating scales for the mini-CEX. *Advances in Health Sciences Education, 14*(5), 655-664.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Holt, Rinehart, & Winston.

Curry, I., Malaktaris, A. L., Lyons, R., Herbert, M. S., & Norman, S. B. (2019). The association between negative trauma-related cognitions and pain-related functional status among veterans with posttraumatic stress disorder and alcohol use disorder. *Journal of Traumatic Stress*, *32*(2), 317–322. https://doi.org/10.1002/jts.22394

Daher, A. M., Ahmad, S. H., Winn, T., & Selamat, M. I. (2015). Impact of rating scale categories on reliability and fit statistics of the Malay Spiritual Well-Being Scale using rasch analysis. *The Malaysian Journal of Medical Sciences: MJMS, 22*(3), 48–55.

Davis, L. W., Schmid, A. A., Daggy, J. K., Yang, Z., O'Connor, C. E., Schalk, N., Do, A. N. L., Maric, D., Lazarick, D., & Knock, H. (2020). Symptoms improve after a yoga program designed for PTSD in a randomized controlled trial with veterans and civilians. *Psychological Trauma: Theory, Research, Practice, and Policy*, *12*(8), 904–912. https://doi.org/10.1037/tra0000564

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430-457. doi:10.1207/S15328007SEM0803_5

Feder, A., Costi, S., Rutter, S. B., Collins, A. B., Govindarajulu, U., Jha, M. K., Horn, S. R., Kautz, M., Corniquel, M., Collins, K. A., Bevilacqua, L., Glasgow, A. M., Brallier, J., Pietrzak, R. H., Murrough, J. W., & Charney, D. S. (2021). A randomized controlled trial of repeated ketamine administration for chronic posttraumatic stress disorder. *American Journal of Psychiatry*, *178*(2), 193–202. https://doi.org/10.1176/appi.ajp.2020.20050596

Fisher, P. W., Lazarov, A., Lowell, A., Arnon, S., Turner, J. B., Bergman, M., Ryba, M., Such, S., Marohasy, C., Zhu, X., Suarez-Jimenez, B., Markowitz, J. C., & Neria, Y. (2021). Equine-assisted therapy for posttraumatic stress disorder among military veterans: An open trial. *The Journal of Clinical Psychiatry*, *82*(5), 36449. https://doi.org/10.4088/JCP.21m14005

Foa, E. B., McLean, C. P., Zang, Y., Zhong, J., Rauch, S., Porter, K., Knowles, K., Powers, M. B., & Kauffman, B. Y. (2016). Psychometric properties of the Posttraumatic Stress Disorder Symptom Scale Interview for DSM–5 (PSSI–5). *Psychological Assessment*, *28*(10), 1159–1165. https://doi.org/10.1037/pas0000259

Fredman, S. J., Macdonald, A., Monson, C. M., Dondanville, K. A., Blount, T. H., Hall-Clark, B. N., Fina, B. A., Mintz, J., Litz, B. T., Young-McCaughan, S., Hancock, A. K., Rhoades, G. K., Yarvis, J. S., Resick, P. A., Roache, J. D., Le, Y., Wachen, J. S., Niles, B. L., McGeary, C. A., … Peterson, A. L. (2020). Intensive, multi-couple group therapy for PTSD: A nonrandomized pilot study with military and veteran dyads. *Behavior Therapy, 51*(5), 700–714. https://doi.org/10.1016/j.beth.2019.10.003

Galovski, T. E., Werner, K. B., Weaver, T. L., Morris, K. L., Dondanville, K. A., Nanney, J., Wamser-Nanney, R., McGlinchey, G., Fortier, C. B., & Iverson, K. M. (2021). Massed cognitive processing therapy for posttraumatic stress disorder in women survivors of intimate partner violence. *Psychological Trauma: Theory, Research, Practice, and Policy*. https://doi.org/10.1037/tra0001100

Gielkens, E. M. J., de Jongh, A., Sobczak, S., Rossi, G., van Minnen, A., Voorendonk, E. M., Rozendaal, L., & van Alphen, S. P. J. (2021). Comparing intensive trauma-focused treatment outcome on PTSD symptom severity in older and younger adults. *Journal of Clinical Medicine*, *10*(6), 1246. https://doi.org/10.3390/jcm10061246

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But… *Communication Methods and Measures, 14*, 1-24. https://doi.org/10.1080/19312458.2020.1718629

Isserles, M., Tendler, A., Roth, Y., Bystritsky, A., Blumberger, D. M., Ward, H., Feifel, D., Viner, L., Duffy, W., Zohar, J., Keller, C. J., Bhati, M. T., Etkin, A., George, M. S., Filipcic, I., Lapidus, K., Casuto, L., Vaishnavi, S., Stein, A., … Ressler, K. J. (2021). Deep transcranial magnetic stimulation combined with brief exposure for posttraumatic

stress disorder: A Prospective multisite randomized trial. *Biological Psychiatry*, *90*(10), 721–728. https://doi.org/10.1016/j.biopsych.2021.04.019

Jeffirs, S. M., Petri, J. M., Camden, A. A., Jackson, B. N., & Weathers, F. W. (2023). Psychometric evaluation of the Multiscale Dissociation Inventory (MDI): A replication and extension in two trauma-exposed samples. *European Journal of Trauma & Dissociation*, *7*(2), 100324. https://doi.org/10.1016/j.ejtd.2023.100324

Kelly, U., Haywood, T., Segell, E., & Higgins, M. (2021). Trauma-sensitive yoga for post-traumatic stress disorder in women veterans who experienced military sexual trauma: Interim results from a randomized controlled trial. *The Journal of Alternative and Complementary Medicine*, *27*(S1), S-45-S-59. https://doi.org/10.1089/acm.2020.0417

Korte, K. J., Capron, D. W., Zvolensky, M., & Schmidt, N. B. (2013). The Fagerström test for nicotine dependence: Do revisions in the item scoring enhance the psychometric properties? *Addictive Behaviors*, *38*(3), 1757–1763. https://doi.org/10.1016/j.addbeh.2012.10.013

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*(2), 363-374. https://www.jstor.org/stable/2529786

Lang, A. J., Malaktaris, A. L., Casmar, P., Baca, S. A., Golshan, S., Harrison, T., & Negi, L. (2019). Compassion meditation for posttraumatic stress disorder in veterans: A randomized proof of concept study. *Journal of Traumatic Stress*, *32*(2), 299–309. https://doi.org/10.1002/jts.22397

Lee, D. J., Thompson-Hollands, J., Strage, M. F., Marx, B. P., Unger, W., Beck, J. G., & Sloan, D. M. (2020). Latent factor structure and construct validity of the Cognitive Emotion

Regulation Questionnaire–Short Form among two PTSD samples. *Assessment, 27*(3), 423–431. https://doi.org/10.1177/1073191118791301

Lim, S. L., Lin, X. H., & Daniels, L. (2016). Seven-point subjective global assessment is more time sensitive than conventional subjective global assessment in detecting nutrition changes. *Journal of Parenteral and Enteral Nutrition*, *40*(7), 966–972. https://doi.org/10.1177/0148607115579938

Maples-Keller, J. L., Hyatt, C. S., Sleep, C. E., Stevens, J. S., Fenlon, E. E., Jovanovic, T., Rothbaum, B. O., Ressler, K. J., Carter, S., Bradley, B., Fani, N., Powers, A., & Michopoulos, V. (2021). DSM–5 alternative model for personality disorders trait domains and PTSD symptoms in a sample of highly traumatized African American women and a prospective sample of trauma center patients. *Personality Disorders: Theory, Research, and Treatment*, *12*(6), 491–502. https://doi.org/10.1037/per0000477

Martin, C. E., Pukay-Martin, N. D., Blain, R. C., Dutton-Cox, C., & Chard, K. M. (2021). Suicidal ideation in a veterans affairs residential posttraumatic stress disorder treatment setting: The roles of thwarted belongingness and perceived burdensomeness. *Journal of Traumatic Stress*, *34*(6), 1188–1198. https://doi.org/10.1002/jts.22540

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30-46. http://dx.doi.org/10.1037/1082-989X.1.1.30

McLean, C. P., Foa, E. B., Dondanville, K. A., Haddock, C. K., Miller, M. L., Rauch, S. A. M., Yarvis, J. S., Wright, E. C., Hall-Clark, B. N., Fina, B. A., Litz, B. T., Mintz, J., Young-McCaughan, S., & Peterson, A. L. (2021). The effects of web-prolonged exposure among military personnel and veterans with posttraumatic stress disorder. *Psychological*

Trauma: Theory, Research, Practice, and Policy, *13*(6), 621–631. https://doi.org/10.1037/tra0000978

Meyer, E. C., Walser, R., Hermann, B., La Bash, H., DeBeer, B. B., Morissette, S. B., Kimbrel, N. A., Kwok, O. M., Batten, S. V., & Schnurr, P. P. (2018). Acceptance and commitment therapy for co-occurring posttraumatic stress disorder and alcohol use disorders in veterans: Pilot treatment outcomes. *Journal of Traumatic Stress*, *31*(5), 781–789. https://doi.org/10.1002/jts.22322

Miller, K. E., Davis, J. L., & Rhudy, J. L. (2018). Pilot study: Brief posttrauma nightmare treatment for persons with bipolar disorder. *Dreaming*, *28*(2), 150–161. https://doi.org/10.1037/drm0000082

Mitchell, J. M., Bogenschutz, M., Lilienstein, A., Harrison, C., Kleiman, S., Parker-Guilbert, K., Ot'alora G., M., Garas, W., Paleos, C., Gorman, I., Nicholas, C., Mithoefer, M., Carlin, S., Poulter, B., Mithoefer, A., Quevedo, S., Wells, G., Klaire, S. S., van der Kolk, B., … Doblin, R. (2021). MDMA-assisted therapy for severe PTSD: A randomized, double-blind, placebo-controlled phase 3 study. *Nature Medicine*, *27*(6), 1025–1033. https://doi.org/10.1038/s41591-021-01336-3

Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory (PAI)*. Psychological Assessment Resources.

Müller-Engelmann, M., Schnyder, U., Dittmann, C., Priebe, K., Bohus, M., Thome, J., Fydrich, T., Pfaltz, M. C., & Steil, R. (2020). Psychometric properties and factor structure of the German version of the Clinician-Administered PTSD Scale for DSM-5. *Assessment*, *27*(6), 1128–1138. https://doi.org/10.1177/1073191118774840

Muthén, L. K., & Muthén, B. (2017). *Mplus user's guide: Statistical analysis with latent variables, user's guide*. Muthén & Muthén.

Norman, S. B., Trim, R., Haller, M., Davis, B. C., Myers, U. S., Colvonen, P. J., Blanes, E., Lyons, R., Siegel, E. Y., Angkaw, A. C., Norman, G. J., & Mayes, T. (2019). Efficacy of integrated exposure therapy vs integrated coping skills therapy for comorbid posttraumatic stress disorder and alcohol use disorder: A randomized clinical trial. *JAMA Psychiatry*, *76*(8), 791. https://doi.org/10.1001/jamapsychiatry.2019.0638

Nursey, J., Sbisa, A., Knight, H., Ralph, N., Cowlishaw, S., Forbes, D., O'Donnell, M., Hinton, M., Cooper, J., Hopwood, M., McFarlane, A., Herring, S., & Fitzgerald, P. (2020). Exploring theta burst stimulation for post-traumatic stress disorder in Australian veterans—A pilot study. *Military Medicine*, *185*(9–10), e1770–e1778. https://doi.org/10.1093/milmed/usaa149

Oaster, T. R. F. (1989). Number of alternatives per choice point and stability of likert-type scales. *Perceptual and Motor Skills*, *68*(2), 549–550. https://doi.org/10.2466/pms.1989.68.2.549

Oprel, D. A. C., Hoeboer, C. M., Schoorl, M., Kleine, R. A. de, Cloitre, M., Wigard, I. G., van Minnen, A., & van der Does, W. (2021). Effect of prolonged exposure, intensified prolonged exposure and STAIR+prolonged exposure in patients with PTSD related to childhood abuse: A randomized controlled trial. *European Journal of Psychotraumatology*, *12*(1), 1851511. https://doi.org/10.1080/20008198.2020.1851511

Ostacher, M. J., Fischer, E., Bowen, E. R., Lyu, J., Robbins, D. J., & Suppes, T. (2021). Investigation of a capnometry guided respiratory intervention in the treatment of

posttraumatic stress disorder. *Applied Psychophysiology and Biofeedback*, *46*(4), 367-376.

Pearse, N. (2011). Deciding on the scale granularity of response categories of Likert type scales: The case of a 21-point scale. *The Electronic Journal of Business Research Methods*, *9*(2), 14.

Petri, J. M., Weathers, F. W., Witte, T. K., & Silverstein, M. W. (2020). The Detailed Assessment of Posttraumatic Stress–Second Edition (DAPS-2): Initial psychometric evaluation in an MTurk-recruited, trauma-exposed community sample. *Assessment, 27*(6), 1116–1127. https://doi.org/10.1177/1073191119880963

Philip, N. S., Barredo, J., Aiken, E., Larson, V., Jones, R. N., Shea, M. T., Greenberg, B. D., & van't Wout-Frank, M. (2019). Theta-burst transcranial magnetic stimulation for posttraumatic stress disorder. *American Journal of Psychiatry*, *176*(11), 939–948. https://doi.org/10.1176/appi.ajp.2019.18101160

Powers, A., Fani, N., Carter, S., Cross, D., Cloitre, M., & Bradley, B. (2017). Differential predictors of DSM-5 PTSD and ICD-11 complex PTSD among African American women. *European Journal of Psychotraumatology*, *8*(1), 1338914. https://doi.org/10.1080/20008198.2017.1338914

Poythress, N. G., Edens, J. F., Skeem, J. L., Lilienfeld, S. O., Douglas, K. S., Frick, P. J., Patrick, C. J., Epstein, M., & Wang, T. (2010). Identifying subtypes among offenders with antisocial personality disorder: a cluster-analytic study. *Journal of Abnormal Psychology, 119*(2), 389–400. https://doi.org/10.1037/a0018611

Prins, A., Bovin, M. J., Smolenski, D. J., Marx, B. P., Kimerling, R., Jenkins-Guarnieri, M. A., Kaloupek, D. G., Scnurr, P. P., Kaiser, A. P., Leyva, Y. E., & Tiet, Q. Q. (2016). The

primary care PTSD screen for DSM-5 (PC-PTSD-5): Development and evaluation within a veteran primary care sample. *Journal of General Internal Medicine, 31*(10), 1206-1211. https://doi.org/10.1007/s11606-016-3703-5

Rae Olmsted, K. L., Bartoszek, M., Mulvaney, S., McLean, B., Turabi, A., Young, R., Kim, E., Vandermaas-Peeler, R., Morgan, J. K., Constantinescu, O., Kane, S., Nguyen, C., Hirsch, S., Munoz, B., Wallace, D., Croxford, J., Lynch, J. H., White, R., & Walters, B. B. (2020). Effect of stellate ganglion block treatment on posttraumatic stress disorder symptoms: A randomized clinical trial. *JAMA Psychiatry*, *77*(2), 130–138. https://doi.org/10.1001/jamapsychiatry.2019.3474

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59 Suppl 20*, 22-33;quiz 34-57.

Shiroma, P. R., Thuras, P., Wels, J., Erbes, C., Kehle-Forbes, S., & Polusny, M. (2020). A proof-of-concept study of subanesthetic intravenous ketamine combined with prolonged exposure therapy among veterans with posttraumatic stress disorder. *The Journal of Clinical Psychiatry*, *81*(6), 10118. https://doi.org/10.4088/JCP.20l13406

Shor, R., Lee, D. J., Thompson-Hollands, J., & Sloan, D. M. (2021). Psychometric properties of the Posttraumatic Avoidance Behaviour Questionnaire among a treatment-seeking adult sample. *Psychological Trauma: Theory, Research, Practice, and Policy*. https://doi.org/10.1037/tra0001060

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420-428. https://doi.org/10.1037//0033-2909.86.2.420

Sippel, L. M., King, C. E., Wahlquist, A. E., & Flanagan, J. C. (2020). A preliminary examination of endogenous peripheral oxytocin in a pilot randomized clinical trial of oxytocin-enhanced psychotherapy for posttraumatic stress disorder. *Journal of Clinical Psychopharmacology*, *40*(4), 401–404. https://doi.org/10.1097/JCP.0000000000001226

Sloan, D. M., Unger, W., Lee, D. J., & Beck, J. G. (2018). A randomized controlled trial of group cognitive behavioral treatment for veterans diagnosed with chronic posttraumatic stress disorder. *Journal of Traumatic Stress*, *31*(6), 886–898. https://doi.org/10.1002/jts.22338

Soder, H. E., Wardle, M. C., Schmitz, J. M., Lane, S. D., Green, C., & Vujanovic, A. A. (2019). Baseline resting heart rate variability predicts post-traumatic stress disorder treatment outcomes in adults with co-occurring substance use disorders and post-traumatic stress. *Psychophysiology*, *56*(8), e13377. https://doi.org/10.1111/psyp.13377

Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology*, *7*(6), 456–461. https://doi.org/10.1037/h0074469

Tang, V. M., Trought, K., Gicas, K. M., Kozak, M., Josselyn, S. A., Daskalakis, Z. J., Blumberger, D. M., Voineskos, D., Knyahnytska, Y., Pasricha, S., Chung, Y., Zhou, Y., Isserles, M., & Wong, A. H. C. (2021). Electroconvulsive therapy with a memory reactivation intervention for post-traumatic stress disorder: A randomized controlled trial. *Brain Stimulation*, *14*(3), 635–642. https://doi.org/10.1016/j.brs.2021.03.015

Thompson-Hollands, J., Lee, D. J., & Sloan, D. M. (2021). The use of a brief family intervention to reduce dropout among veterans in individual trauma-focused treatment: A randomized controlled trial. *Journal of Traumatic Stress*, *34*(4), 829–839. https://doi.org/10.1002/jts.22680

Trottier, K., Monson, C. M., Wonderlich, S. A., & Olmsted, M. P. (2017). Initial findings from Project Recover: Overcoming co-occurring eating disorders and posttraumatic stress disorder through integrated treatment. *Journal of Traumatic Stress*, *30*(2), 173–177. https://doi.org/10.1002/jts.22176

van Minnen, A., van Dalen, B., Voorendonk, E. M., Wagenmans, A., & de Jongh, A. (2020). The effects of symptom overreporting on PTSD treatment outcome. *European Journal of Psychotraumatology*, *11*(1), 1794729. https://doi.org/10.1080/20008198.2020.1794729

van Gelderen, M. J., Nijdam, M. J., Haagen, J. F. G., & Vermetten, E. (2020). Interactive motion-assisted exposure therapy for veterans with treatment-resistant posttraumatic stress disorder: A randomized controlled trial. *Psychotherapy and Psychosomatics*, *89*(4), 215–227. https://doi.org/10.1159/000505977

Vogt, D. S., Shipherd, J. C., & Resick, P. A. (2012). Posttraumatic Maladaptive Beliefs Scale: Evolution of the Personal Beliefs and Reactions scale. *Assessment*, *19*(3), 308–317. https://doi.org/10.1177/1073191110376161

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013). *The Life Events Checklist for DSM-5 (LEC-5) - Extended*. Retrieved from the National Center for PTSD website: www.ptsd.va.gov

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., Keane, T. M., & Marx, B. P. (2018). The Clinician-Administered PTSD Scale for DSM–5

(CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment, 30*(3), 383–395. https://doi.org/10.1037/pas0000486

Weathers, F. W., Keane, T. M., & Davidson, J. R. T. (2001). Clinician-Administered PTSD scale: A review of the first ten years of research. *Depression and Anxiety*, *13*(3), 132–156. https://doi.org/10.1002/da.1029

Weathers, F. W., Ruscio, A. M., & Keane, T. M. (1999). Psychometric properties of nine scoring rules for the Clinician-Administered Posttraumatic Stress Disorder Scale. *Psychological Assessment, 11*(2), 124-133.

Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, *64*(6), 956–972. https://doi.org/10.1177/0013164404268674

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology, 84*, 608–618. doi:10.1037/0022-3514.84.3.608

Zhu, X., Suarez-Jimenez, B., Zilcha-Mano, S., Lazarov, A., Arnon, S., Lowell, A. L., Bergman, M., Ryba, M., Hamilton, A. J., Hamilton, J. F., Turner, J. B., Markowitz, J. C., Fisher, P. W., & Neria, Y. (2021). Neural changes following equine-assisted therapy for posttraumatic stress disorder: A longitudinal multimodal imaging study. *Human Brain Mapping*, *42*(6), 1930–1939. https://doi.org/10.1002/hbm.25360

Appendix

Table 1

Summary of CAPS-5 Findings from Treatment Outcome Studies

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
Psychosocial treatment studies					
1. van Gelderen et al. (2020)	Veterans (n=43)	RCT, assessments blinded	1. 3MDR (virtual reality, EMDR, and walking on treadmill while interacting with images) 2. Non-specific treatment	6 weekly sessions (followed by optional 10 weekly sessions)	1. Significant reduction in CAPS-5 total score for 3MDR condition (16%) compared to NST (6%) at post-treatment 2. Significant between-groups effect, with 3MDR condition showing larger reductions in CAPS-5 total score at post-treatment
2. Bisson et al. (2020)	Veterans (n=42)	RCT, assessments blinded	1. Immediate 3MDR 2. Delayed 3MDR	9 weekly sessions	1. Significant reduction in CAPS-5 total score for immediate 3MDR condition (37%) compared to waitlist (14%) at week 12 2. Significant between-groups effect, with immediate 3MDR condition showing larger reductions in CAPS-5 total score at post-treatment
3. Meyer et al. (2018)	Veterans with comorbid PTSD & AUD (n=29)	Uncontrolled trial, master's level assessor not part of therapy	1. Acceptance and Commitment Therapy	12 sessions	1. Significant reduction in CAPS-5 total score (17%)

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
4. Thompson-Hollands et al. (2021)	Veterans and family members (n=20 dyads)	RCT, assessments blinded	1. Brief family intervention (BFI) with treatment as usual (TAU) 2. Non-BFI with TAU	2 sessions	1. Significant reduction in CAPS-5 total score for both conditions (BFI = 41%, Non-BFI = 30%) 2. No significant between-groups effect
5. Soder et al. (2019)	Civilians with SUD and comorbid PTSD symptoms (n=37)	Portion of sample non-randomized; assessors not involved with therapy, unclear if blinded to condition	1. CBT for SUD 2. CBT for SUD and CPT for PTSD	12 sessions in 6 weeks	1. Significant reduction in CAPS-5 total score for both conditions (CBT for SUD = 25%, CBT for SUD/CPT for PTSD = 33%)
6. Lang et al. (2019)	Veterans (n=28)	RCT, assessments blinded	1. Compassion meditation (CM) 2. Veteran.Calm (VC)	10 weekly sessions	1. Significant reduction in CAPS-5 total score for both conditions (CM = 43%, VC = 7%) 2. Significant between-groups effect, with CM showing larger reductions in CAPS-5 total score at post-treatment
7. Blain et al. (2021)	Veterans (n=107)	Uncontrolled trial	1. CPT	12 sessions in 6 weeks	1. Significant reduction in CAPS-5 total score (42%)

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
8. Bryan et al. (2018)	Military Personnel and Veterans (n=20)	Uncontrolled trial	1. CPT	12 daily sessions	1. Significant reduction in CAPS-5 total score (53%)
9. Bernardi et al. (2019)	Karen refugees (n=7)	Uncontrolled trial	1. CPT	12 weekly sessions	1. Significant reduction in CAPS-5 total score (45%)
10. Galovski et al. (2021)	Women with intimate partner violence (n=12)	Matched pairs, partial randomization	1. Standard CPT 2. Massed CPT	1. sCPT: 12 sessions once or twice weekly 2. mCPT: 12 sessions in 5 days	1. Significant reduction in CAPS-5 total score for both conditions (sCPT = 41%, mCPT = 53%) 2. No significant between-groups effect
11. Bohus et al. (2020)	Treatment-seeking women with childhood abuse (n=193)	RCT, assessments blinded	1. DBT-PTSD 2. CPT	16 sessions	1. Significant reduction in CAPS-5 total score for both conditions (DBT-PTSD = 49%, CPT = 36%) 2. Significant between-groups effect, with DBT-PTSD showing larger reductions in CAPS-5 total score at post-treatment
12. Carletto et al. (2019)	Breast cancer patients with PTSD (n=30)	Quasi-experimental, blinded assessments	1. EMDR 2. TAU	1. TAU: 4 sessions every other week in 2 months 2. EMDR: 10 sessions in 2-3 months	1. Significant reduction in CAPS-5 total score for both conditions (EMDR= 81%, TAU = 16%) 2. Significant between-groups effect, with EMDR showing larger reductions in CAPS-5 total score at post-treatment

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
13. Assmann et al. (2021)	Treatment-seeking patients with PTSD (n=155), separated by MDD+ and MDD-	RCT, assessments blinded	1. EMDR 2. Imagery rescripting	12 sessions in 6 weeks	1. Significant reduction in CAPS-5 total score for both conditions, and both samples (MDD- [IR= 36%, EMDR = 52%], MDD+ [IR = 36%, EMDR = 39%]) 2. Patients without MDD showed greater reductions when treated with EMDR than patients with comorbid MDD
14. Gielkens et al. (2021)	Treatment seeking patients with PTSD (n=124)	Uncontrolled trial	1. EMDR, PE, physical activity, and group psychoeducation	8 days	1. Significant reduction in CAPS-5 total score (61%)
15. Fisher et al. (2021)	Treatment-seeking veterans (n=63)	Uncontrolled trial	1. EAT	8 sessions	1. Significant reduction in CAPS-5 total score (30%)
16. Miller et al. (2018)	Treatment seeking patients with Bipolar I or Bipolar II who reported a Criterion A traumatic event and nightmares (n=7)	Uncontrolled trial, assessments conducted by research assistants	1. Exposure, relaxation, and rescripting therapy supplemented with psychoeducation about bipolar disorder's effect on sleep	5 weekly sessions	1. Significant reduction in CAPS-5 total score (55%)

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
17. Sloan et al. (2018)	Male veterans (n=198)	RCT, assessments blinded	1. Group CBT 2. Group person-centered treatment	14 sessions in 16 weeks	1. Significant reduction in CAPS-5 total score for both conditions (GCBT= 18%, GPCT = 13%) 2. No significant between-groups effect
18. Trottier et al. (2017)	Patients with co-occurring ED and PTSD (n=11)	Uncontrolled trial	1. Integrated CPT with supplemental ED interventions	16 sessions	1. Significant reduction in CAPS-5 total score (73%)
19. Norman et al. (2019)	Treatment-seeking veterans with comorbid AUD & PTSD (n=119)	RCT, assessments blinded	1. Integrated PE 2. Integrated Coping Skills	12 sessions	1. Significant reduction in CAPS-5 total score for both conditions (iPE = 40%, iCS = 22%) 2. Significant between-groups effect, with iPE showing larger reductions in CAPS-5 total score at post-treatment
20. Fredman et al. (2020)	Veterans and partners (n=24 dyads)	Uncontrolled trial, assessors not involved in therapy	1. Intensive cognitive-behavioral conjoint therapy	1 weekend	1. Significant reduction in CAPS-5 total score (22%)

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
21. Oprel et al. (2021)	Outpatients with PTSD and childhood abuse (n=150)	RCT, assessments blinded	1. PE 2. Intensive PE 3. Skills Training in Affective and Interpersonal Regulation + PE (STAIR+PE)	1. PE: 16 weekly sessions 2. iPE: 12 sessions in 4 weeks (followed by 2 follow-up sessions) 3. STAIR+PE: 16 weekly sessions	1. Significant reduction in CAPS-5 total score for all conditions at post-treatment (PE = 57%, iPE = 35%; STAIR+PE = 51%) 2. Between-groups effects at post-treatment not examined due to comparisons being made at the 1-year follow up
22. van Minnen et al. (2020)	Patients with PTSD (n=205), separated by over reporters and non-over reporters	Uncontrolled trial	1. PE and EMDR	8 sessions in 2 weeks	1.Significant reduction in CAPS-5 total score for both samples at post-treatment (Over reporters = 46%; Non-over reporters = 55%) 2. No significant between-groups differences
23. McLean et al. (2021)	Treatment-seeking military personnel (n=40)	RCT, assessments blinded	1. Web-version of PE 2. In-person present-centered therapy (PCT)	10 sessions in 8 weeks	1. Significant reduction in CAPS-5 total score for both conditions (Pre- and post-treatment scores not disclosed) 2. No significant between-groups effect

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
24. Andersen et al. (2021)	Community members with neck trauma (n=104)	RCT, assessments blinded	1. TF-CBT and exercise 2. Supportive therapy	16 weekly sessions	1. Significant reduction in CAPS-5 total score for both conditions (TF-CBT and exercise = 52%, ST = 44%) 2. No significant between-groups effect
25. Kelly et al. (2021)	Female veterans (n=152)	RCT, assessments not blinded	1. Trauma-sensitive yoga 2. CPT	1. Yoga: 10 weekly sessions 2. CPT: 12 weekly sessions	1. Significant reduction in CAPS-5 total score for both conditions (Yoga = 43%, ST = 25%) 2. No significant between-groups effect
26. Davis et al. (2020)	Community members with PTSD (n=209)	RCT, assessments blinded	1. Wellness lifestyle program (WLP) 2. Holistic yoga program (WYP)	16 weekly sessions	1. Significant reduction in CAPS-5 total score for HYP condition (30%) compared to WLP condition (14%)
27. Sloan et al. (2018)	Community (n=126)	RCT, assessments blinded	1. Written exposure therapy (WET) 2. CPT	1. WET: 5 sessions 2. CPT: 12 sessions once or twice a week	1. Significant reduction in CAPS-5 total score for both conditions (WET = 26%, ST = 33%) at 12 weeks 2. No significant between-groups effect

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
Pharmacological treatment studies					
28. Mitchell et al. (2021)	Patients with severe PTSD (n=90)	RCT, assessments blinded	1. Therapy supplemented with MDMA 2. Placebo	3, 8-hour sessions	1. Significant reduction in CAPS-5 total score for both conditions (MDMA = 55%, Placebo = 32%) 2. Significant between-groups effect, with MDMA condition showing larger reductions in CAPS-5 total score at post-treatment
29. Rae Olmsted et al. (2020)	Military service members (n=113)	RCT, assessments blinded	1. Stellate ganglion block 2. Placebo	2 weekly sessions	1. Significant reduction in CAPS-5 total score for both conditions (SGB = 32%, Placebo = 15%) 2. Significant between-groups effect, with SGB condition showing larger reductions in CAPS-5 total score at post-treatment
30. Bonn-Miller et al. (2021)	Community (n=80)	Randomized cross-over trial, assessments blinded	First stage: 1. High THC 2. High CBD 3. THC + CBD 4. Placebo Second stage: 1. High THC 2. High CBD 3. Placebo	1.8 grams a day for 3 weeks	1. Significant reduction in CAPS-5 total score for all conditions in Stage 1 (High THC = 44%, High CBD = 24%, THC + CBD = 22%, Placebo = 35%) 2. No significant between-groups effect in stage 1

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
31. Shiroma et al. (2020)	Veterans (n=9)	Uncontrolled trial, unclear if assessments blinded	1. PE supplemented with ketamine	3 weekly sessions before PE	1. Significant reduction in CAPS-5 total score (69%) 2. Significance went away once Depression (assessed via the MADRS) was controlled for
32. Feder et al. (2021)	Community members with PTSD (n=30)	RCT, assessments blinded	1. Ketamine 2. Midalazolam	6 sessions in 2 weeks	1. Significant reduction in CAPS-5 total score for both conditions (pre- and post-treatment scores not available) 2. Significant between-groups effect, with Ketamine condition showing larger reductions in CAPS-5 total score at post-treatment
33. Sippel et al. (2020)	Community (n=17)	RCT, assessments blinded	1. PE supplemented with Oxytocin 2. Placebo	10 sessions	1. Significant reduction in CAPS-5 total score for both conditions (Oxytocin = 66%, Placebo = 53%) 2. No significant between-groups effect
			Other treatment studies		
34. Nursey et al. (2020)	Veterans (n=8)	Case series, repeated measures	1. Theta-burst stimulation	5 weekly sessions in 4 weeks	1. Significant reduction in CAPS-5 total score (22%)

Table 1 continued.

Authors (year)	Sample(s)	Design	Treatment	Number of sessions/duration	Key findings
35. Philip et al. (2019)	Veterans (n=50)	RCT, assessments blinded	1. Theta-burst stimulation 2. Placebo	10 daily sessions	1. Significant reduction in CAPS-5 total score for both conditions (TBS = 19%, Placebo = 17%) 2. No significant between-groups effect
36. Isserles et al. (2021)	Outpatients (n=125)	RCT, assessments blinded	1. Deep transcranial magnetic stimulation 2. Placebo	12 sessions in 4 weeks	1. Significant reduction in CAPS-5 total score for both conditions (dTMS = 38%, Placebo = 50%) 2. Significant between-groups effect, with Placebo condition showing larger reductions in CAPS-5 total score at post-treatment
37. Tang et al. (2021)	Patients with severe depression or suicidality (n=28)	RCT, assessments blinded	1. ECT with traumatic memory 2. ECT with neutral memory	Number of sessions not indicated	1. Significant reduction in CAPS-5 total score for both conditions (traumatic memory = 37%, neutral memory = 38%)
38. Ostacher et al. (2021)	Community (n=55)	Uncontrolled trial	1. Capnometry guided respiratory intervention	2 sessions daily for 4 weeks	1. Significant reduction in CAPS-5 total score (36%)

Table 2

Sociodemographic Characteristics of Participants

Baseline characteristic	CAPS-5-R test-retest condition	Alternate forms condition	Full sample
		n (%)	
Sex			
Male	8 (27.6)	9 (31.0)	22 (30.1)
Female	21 (72.4)	20 (69.0)	51 (69.9)
Race/Ethnicity			
White	25 (86.2)	25 (86.2)	61 (83.6)
Black/African American	2 (6.9)	2 (6.9)	8 (11.0)
Hispanic/Latine	2 (6.9)	2 (6.9)	6 (6.8)
Multi-racial	2 (6.9)	1 (3.4)	3 (4.1)
American Indian/Alaskan Native	0 (0)	1 (3.4)	1 (1.4)
Sexual and gender identity			
Straight (heterosexual)	24 (82.8)	22 (75.9)	56 (76.7)
Lesbian/gay	2 (6.9)	3 (10.3)	5 (6.8)
Bisexual	2 (6.9)	4 (13.8)	10 (13.7)
Transgender	1 (3.4)	0 (0)	2 (2.7)
Marital status			
Single	5 (17.2)	8 (27.6)	17 (23.3)

Married/partnered	14 (48.3)	18 (62.1)	40 (54.8)
Divorced/separated/widowed	10 (34.5)	3 (10.3)	16 (21.9)
Highest education level			
GED	1 (3.4)	3 (10.3)	5 (6.8)
High school	2 (6.9)	3 (10.3)	5 (6.8)
Some college or associates degree	14 (48.3)	15 (51.7)	37 (50.7)
University or postgraduate degree	12 (41.4)	8 (27.6)	26 (35.6)
Employment status			
Unemployed	11 (37.9)	12 (41.4)	29 (39.7)
Part-time	4 (13.8)	7 (24.1)	12 (16.4)
Full-time	14 (48.3)	9 (31.0)	31 (42.5)
Income level			
< $30,000	7 (24.1)	16 (55.2)	29 (39.7)
$30,000–$59,999	11 (37.9)	3 (10.3)	20 (27.4)
≥ $60,000	18 (62.1)	10 (34.4)	24 (32.9)
Student	4 (13.7)	4 (13.7)	9 (12.3)
Veteran/military service member	2 (6.9)	0 (0)	3 (4.1)
MINI Comorbid Diagnosis[a]			
Major Depressive Disorder	8 (28.6)	5 (17.9)	17 (25.0)
Generalized Anxiety Disorder	10 (35.7)	5 (17.9)	20 (29.4)

Panic Disorder	4 (14.3.0)	4 (14.3)	10 (14.7)
Agoraphobia	5 (17.9)	7 (25.0)	15 (22.1)
Social Anxiety Disorder	3 (10.7)	5 (17.9)	11 (16.2)
Obsessive-Compulsive Disorder	1 (3.6)	1 (3.6)	3 (4.4)
Alcohol Use Disorder	5 (17.9)	2 (7.1)	9 (13.3)
Substance Use Disorder	6 (21.4)	0 (0.0)	8 (11.8)
Bulimia Nervosa	0 (0.0)	1 (3.6)	1 (1.5)
Binge Eating Disorder	2 (7.1)	1 (3.6)	3 (4.4)

Note. MINI = Mini International Neuropsychiatric Interview; n = 29 for both retest conditions; full sample (n = 73) includes participants without retest conditions. Participants were on average 42.6 years old (SD = 13.1), and participant age did not different between retest conditions.
[a]MINI comorbidity data for 5 participants were not collected (ns = 68 and 28 for the full sample and retest conditions, respectively).

Table 3

Descriptive Statistics of Criterion Measures

Variable	*n*	*M*	*SD*	Observed range	α
Clinician-Administered PTSD Scale for *DSM-5* (CAPS-5)	33	26.4	15.8	1 – 53	.93
PTSD Checklist for *DSM-5* Total Score (PCL-5)	71	38.8	18.9	6 – 72	.94
Multiscale Dissociation Inventory (MDI)	71	63.3	24.8	31 – 123	.96
Detailed Assessment for Posttraumatic Stress – 2 (DAPS-2)					
Posttraumatic Stress Total	69	125.5	42.4	46 – 210	.97
Personality Assessment Inventory (PAI) Scales					
Anxiety	70	38.1	15.5	9 – 72	.93
Depression	71	35.8	16.1	5 – 68	.93
Mania	66	26.0	10.7	7 – 49	.84
Anxiety-Related Disorders	72	36.6	12.5	9 – 66	.86
Paranoia	69	30.7	12.5	8 – 61	.91
Schizophrenia Features	68	23.7	10.9	5 – 58	.87
Borderline Features	70	34.0	12.0	11 – 64	.87

Antisocial Features	70	16.0	9.9	1 – 39	.82
Somatic Concerns	66	29.0	15.6	4 – 60	.92
Alcohol Problems	71	5.0	5.7	0 – 27	.81
Drug Problems	71	8.0	7.1	0 – 29	.82

Table 4

Descriptive Statistics of CAPS-5-R Total, Cluster, and Item Scores at Time 1

Scale/Item	*M*	*SD*	Range observed (possible)	α	Ω	CEP %
Total	53.3	30.1	7 – 124 (0 – 200)	.89	.90	--
Reexperiencing	12.7	7.9	0 – 30 (0 – 50)	.75	.76	--
B1	3.7	2.2	0 – 9	--	--	72.7
B2	2.4	2.8	0 – 9	--	--	41.8
B3	0.7	1.6	0 – 7	--	--	16.4
B4	3.2	2.1	0 – 10	--	--	69.1
B5	2.7	2.4	0 – 8	--	--	54.5
Avoidance	6.5	4.8	0 – 18 (0 – 20)	.61	.61	--
C1	3.5	3.0	0 – 9	--	--	56.4
C2	3.1	2.7	0 – 9	--	--	56.4
Negative Alterations in Cognition and Mood	18.8	12.9	0 – 48 (0 – 70)	.77	.80	--
D1	1.0	2.3	0 – 10	--	--	18.2
D2	3.3	3.3	0 – 10	--	--	52.7

D3	3.3	2.8	0 – 9	--	--	61.8
D4	3.3	2.7	0 – 9	--	--	61.8
D5	1.7	2.5	0 – 8	--	--	34.5
D6	3.3	3.3	0 – 9	--	--	50.9
D7	2.9	2.8	0 – 9	--	--	56.4
Alterations in Arousal and Reactivity	15.2	10.4	0 – 37 (0 – 60)	.73	.79	--
E1	1.7	1.9	0 – 6	--	--	34.5
E2	0.2	0.9	0 – 4	--	--	5.5
E3	4.4	3.3	0 – 10	--	--	70.9
E4	1.9	2.1	0 – 6	--	--	43.6
E5	2.6	3.0	0 – 8	--	--	45.5
E6	4.4	3.7	0 – 10	--	--	60.0
Depersonalization	0.4	1.0	0 – 4	--	--	9.1
Derealization	0.5	1.2	0 – 6	--	--	10.9

Note. n = 55; CEP % = clinical elevation prevalence (percentage of respondents with item severity scores greater than or equal to 3); possible range for individual CAPS-5-R items is 0 to 10.

Table 5

Reliability Coefficients of CAPS-5-R Total, Cluster, and Item Scores

	Test-Retest ICC (95% C. I.)		Interrater ICC (95% C. I.)		Alternate Forms ICC (95% C. I.)	
Scale/Item	CAPS-5-R	CAPS-5-R with CAPS-5 scoring	CAPS-5-R	CAPS-5-R with CAPS-5 scoring	CAPS-5-R with original scoring	CAPS-5-R with CAPS-5 scoring
Total	.86 (.73, .93)	.87 (.74, .94)	.98 (.96, .99)	.98 (.96, .99)	.38 (.02, .65)	.94 (.88, .97)
RE	.81 (.64, .91)	.77 (.57, .89)	.96 (.92, .98)	.97 (.93, .98)	.36 (.00, .64)	.78 (.59, .89)
B1	.63 (.36, 81)	.49 (.16, .72)	.85 (.71, .93)	.81 (.64, .91)	.18 (-.19, .51)	.57 (.27, .77)
B2	.76 (.56, .88)	.71 (.47, .85)	.99 (.97, .99)	.98 (.96, .99)	.44 (.09, .69)	.73 (.51, .86)
B3	.56 (.26, .77)	.66 (.40, .82)	.83 (.68, .92)	.96 (.93, .98)	.49 (.16, .72)	.62 (.34, .80)
B4	.43 (.08, .68)	.12 (-.25, .46)	.90 (.81, .95)	.86 (.73, .93)	.29 (-.07, .59)	.56 (.26, .77)
B5	.49 (.16, .72)	.57 (.26, .77)	.95 (.91, .98)	.96 (.91, .98)	.34 (-.02, .62)	.63 (.35, .81)
AV	.66 (.40, .83)	.75 (.53, .87)	.92 (.85, .96)	.91 (.82, .96)	.41 (.06, .67)	.86 (.72, .93)
C1	.59 (.29, .78)	.49 (.16, .72)	.90 (.80, .95)	.89 (.77, .94)	.33 (-.03, .62)	.71 (.48, .85)
C2	.50 (.17, .73)	.61 (.32, .80)	.95 (.90, .98)	.92 (.83, .96)	.48 (.15, .72)	.81 (.64, .91)

NACM	.79 (.61, .90)	.78 (.59, .89)	.97 (.93, .98)	.95 (.90, .98)	.46 (.12, .70)	.86 (.73, .93)
D1	.71 (.47, .85)	.80 (.62, .90)	.84 (.69, .92)	.82 (.66, .91)	.43 (.09, .68)	.65 (.38, .82)
D2	.64 (.37, .81)	.64 (.37, .81)	.96 (.91, .98)	.96 (.93, .98)	.30 (-.07, .59)	.74 (.52, .87)
D3	.57 (.30, .79)	.61 (.32, .79)	.91 (.83, .96)	.92 (.83, .96)	.27 (-.10, .57)	.63 (.35, .81)
D4	.30 (-.06, .60)	.27 (-.09, .58)	.89 (.78, .95)	.87 (.75, .94)	.23 (-.14, .54)	.51 (.18, .73)
D5	.58 (.29, .78)	.64 (.37, .81)	.93 (.86, .97)	.85 (.67, .92)	.60 (.31, .79)	.68 (.42, .83)
D6	.58 (.28, .78)	.43 (.09, .69)	.96 (.92, .98)	.96 (.92, .98)	.48 (.14, .71)	.77 (.57, .89)
D7	.66 (.40, .83)	.70 (.45, .85)	.92 (.83, .96)	.93 (.85, .96)	.44 (.09, .69)	.70 (.56, .85)
AAR	.82 (.65, .91)	.77 (.57, .89)	.98 (.96, .99)	.98 (.95, .99)	.25 (-.12, .56)	.82 (.65, .91)
E1	.62 (.34, .80)	.59 (.30, .78)	.87 (.74, .94)	.92 (.83, .96)	.25 (-.12, .56)	.39 (.03, .65)
E2	.41 (.06, .67)	.44 (.10, .69)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
E3	.65 (.38, .82)	.67 (.42, .83)	.92 (.85, .96)	.92 (.83, .96)	.09 (-.28, .44)	.76 (.55, .88)
E4	.64 (.37, .81)	.61 (.32, .79)	.87 (.74, .93)	.93 (.85, .97)	.36 (-.00, .63)	.64 (.37, .81)
E5	.83 (.67, .92)	.79 (.60, .90)	.96 (.92, .98)	.96 (.92, .98)	.50 (.17, .73)	.89 (.78, .95)
E6	.76 (.56, .88)	.71 (.47, .85)	.96 (.93, .98)	.96 (.91, .98)	.30 (-.07, .59)	.50 (.18, .73)
Depers	.63 (.35, .81)	.50 (.17, .73)	.96 (.91, .98)	.97 (.95, .99)	.61 (.32, .80)	.71 (.48, .85)

Dereal	.55 (.24, .76)	.32 (-.05, .60)	.63 (.35, .80)	.72 (.48, .86)	.63 (.35, .81)	.82 (.65, .91)

Note. n = 29; RE = Reexperiencing cluster; AV = Avoidance cluster; NACM = Negative Alterations in Cognition and Mood cluster; AAR = Alterations in Arousal and Reactivity cluster; Depers = Depersonalization; Dereal = Derealization.

Table 6

Predicted and Observed Correlations Between CAPS-5-R, CAPS-5, and Criterion Measures

	Observed Correlations (*r*)					Predicted Correlations (*r*)
	Time 1 only		Alternate Forms			
Variable	CAPS-5-R	CAPS-5-R with CAPS-5 scoring	CAPS-5-R	CAPS-5-R with CAPS-5 scoring	CAPS-5	
PCL-5	.72**	.66**	.80**	.76**	.80**	.70
DAPS-2-PTS-T	.69**	.66**	.84**	.80**	.85**	.70
PAI- ARDT	.72**	.69**	.82**	.80**	.85**	.70
MDI	.56**	.54**	.70**	.67**	.73**	.60
PAI-DEP	.58**	.55**	.70**	.69**	.74**	.60
PAI-ANX	.46**	.42**	.70**	.67**	.66**	.50
PAI-BOR	.49**	.47**	.55**	.55**	.61**	.50
PAI-SCZ	.55**	.53**	.76**	.75**	.78**	.40
PAI-PAR	.63**	.61**	.59**	.61**	.69**	.40
PAI-SOM	.33*	.32*	.46*	.44*	.45*	.40
PAI-ALC	.14	.12	.13	.08	.20	.20
PAI-DRG	.24	.23	-.18	-.28	-.08	.20
PAI-ANT	.25	.25	.12	.05	.22	.10
PAI-MAN	.33*	.34*	.24	.26	.31	.10

Note. n = 49-55 for Time 1 only; n = 27-29 for Alternate Forms, which included only those participants in the CAPS-5 vs. CAPS-5-R retest condition; CAPS-5-R = Revised Clinician-Administered PTSD Scale for *DSM-5*; CAPS-5= Clinician-Administered PTSD Scale for *DSM-5*; PCL-5 = PTSD Checklist for *DSM-5*; DAPS-2-PTS-T = Detailed Assessment of Posttraumatic Stress (2nd Ed) – Posttraumatic Stress Total Scale; PAI = Personality Assessment Inventory; ARDT = Anxiety Related Disorders-Traumatic Stress subscale; DEP = Depression scale; ANX = Anxiety scale; BOR = Borderline scale; SCZ = Schizophrenia scale; PAR = Paranoia Scale; SOM = Somatization scale; ALC = Alcohol Problems scale; DRG = Drug Problems scale; ANT = Antisocial Scale; MAN = Mania scale; MDI = Multiscale Dissociation Inventory. Lambdas are available upon request from the first author.

* $p < .05$

** $p < .01$

Table 7

Effect Size Statistics ($r_{alerting\text{-}CV}$ and $r_{contrast\text{-}CV}$) for CAPS-5-R and CAPS-5

Quantity	Time 1 only		Alternate Forms		
	CAPS-5-R	CAPS-5-R with CAPS-5 scoring	CAPS-5-R	CAPS-5-R with CAPS-5 scoring	CAPS-5
$r_{alerting\text{-}CV}$	.88	.86	.90	.88	.89
$r_{contrast\text{-}CV}$	.81	.76	.98	.97	.97
95% CI	[.69, .89]	[.62, .86]	[.95, .99]	[.94, .99]	[.93, .99]
$Z_{contrast}$	7.20	6.47	8.87	8.64	8.38
$t_{contrast}$	9.62	8.15	23.29	21.42	19.53
$p_{contrast}$	<.001	<.001	<.001	<.001	<.001

Note. The following harmonic means were used for computation: n = 52 for Time 1 only and n = 28 for Alternate Forms. CAPS-5-R = Revised Clinician-Administered PTSD Scale for *DSM-5*; CAPS-5= Clinician-Administered PTSD Scale for *DSM-5*.

Table 8

Results of the Rater Feedback Survey

Evaluation of the CAPS-5-R across different domains	Much worse	Somewhat worse	About the same	Somewhat better	Much better
			n (%)		
Ease of administration	0	0	1 (20)	0	4 (80)
Ease of scoring	0	0	1 (20)	1 (20)	3 (60)
Clarity of prompts	0	0	0	2 (40)	3 (60)
Clarity of scoring instructions	0	0	0	1 (20)	4 (80)
Length of administration	0	0	4 (80)	0	1 (20)
Comprehensiveness of item content	0	0	0	2 (40)	3 (60)
Overall participant burden	0	0	1 (20)	2 (40)	2 (40)
Overall rater burden	0	0	0	3 (60)	2 (40)

Table 7 cont.

Results of the Rater Feedback Survey

Receptiveness to CAPS-5-R changes	Very much worsened the instrument	Somewhat worsened the instrument	Neither improved nor worsened the instrument	Somewhat improved the instrument	Very much improved the instrument
	n (%)				
Addition of the frequency rating card	0	0	0	1 (20)	4 (80)
Splitting of item D3	0	0	0	2 (40)	3 (60)
Change in D1 frequency assessment	0	0	0	3 (60)	2 (40)
Rewording/addition of item prompts	0	0	0	1 (20)	4 (80)
Initial instructions for participant	0	0	0	1 (20)	4 (80)
Expansion of the rating scale	0	0	0	1 (20)	4 (80)
Scoring guidelines	0	0	0	1 (20)	4 (80)
Rating grid on each page	0	0	0	0	5 (100)

Figure 1

CAPS-5 Scoring Grid

CAPS-5 Past Month Scoring Rules			
INTENSITY	***FREQUENCY (# of times or %)***		***SEVERITY***
Minimal	*1x/month or more*	*1-100%*	*1 = Mild / subthreshold*
Clearly Present	*1x/month only*	*1-19%*	*1 = Mild / subthreshold*
	2x/month or more	*20-100%[a]*	*2 = Moderate / threshold*
Pronounced	*At least 1x/month but less than 2x/week*	*1-49%[a]*	*2 = Moderate / threshold*
	2x/week or more	*50-100%[b]*	*3 = Severe / markedly elevated*
Extreme	*1x/month only*	*1-19%[a]*	*2 = Moderate / threshold*
	At least 2x/month but not daily or almost every day	*20-79%[b]*	*3 = Severe / markedly elevated*
	Daily/almost every day	*80-100%[c]*	*4 = Extreme / incapacitating*

Figure 2

CAPS-5-R Scale Anchors and Scoring Grid

Intensity	Frequency (# of times)
0 = Absent	0 = Absent
1 = Minimal	1 = 1x/month
2 = Clearly Present	2 = 2-3x/month
3 = Clearly Present/Pronounced	3 = 1x/week (4-7x/month)
4 = Pronounced	4 = 2-3x/week (8-15x/month)
5 = Pronounced/Extreme	5 = 4-5x/week (16-23x/month)
6 = Extreme	6 = 6-7x/week (24-30x/month)

	FREQUENCY (# times or % of time)						
	0	1x	2-3x	1x/week (4-7x)	2-3x/week (8-15x)	4-5x/week (16-23x)	6-7x/week (24-30x)
INTENSITY	0	< 10%	10-19%	20-39%	40-59%	60-79%	80-100%
Absent	0	0	0	0	0	0	0
Minimal	0	1	1	1	1	2	2
Clearly Present	0	2	3	3	4	4	5
Clearly Present/ Pronounced	0	2	3	4	5	6	6
Pronounced	0	3	4	5	6	7	7
Pronounced/ Extreme	0	4	5	6	6	7	8
Extreme	0	5	6	7	8	9	10

Figure 3

CAPS-5-R Severity Rating Anchors

Absent	Mild / subthreshold		Moderate / threshold			Severe / markedly elevated			Extreme / incapacitating	
0	1	2	3	4	5	6	7	8	9	10