EUROPEAN JOURNAL OF
PSYCHO-
TRAUMATOLOGY
THE OFFICIAL JOURNAL OF THE EUROPEAN SOCIETY FOR TRAUMATIC STRESS STUDIES

Taylor & Francis
Taylor & Francis Group

CLINICAL RESEARCH ARTICLE

A comparison of the CAPS-5 and PCL-5 to assess PTSD in military and veteran treatment-seeking samples

Patricia A. Resick [a], Casey L. Straud [b,c,d], Jennifer Schuster Wachen [e,f], Stefanie T. LoSavio [b], Alan L. Peterson [b,c,d], Donald D. McGeary [b,d], Stacey Young-McCaughan [b,d], Daniel J. Taylor [g] and Jim Mintz [b,d] for the STRONG STAR Consortium and the Consortium to Alleviate PTSD

aDepartment of Psychiatry and Behavioral Sciences, Duke Health, Durham, NC, USA; bDepartment of Psychiatry and Behavioral Sciences, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA; cDepartment of Psychology, University of Texas at San Antonio, San Antonio, TX, USA; dSouth Texas Veterans Health Care System, San Antonio, TX, USA; eWomen's Health Sciences Division, National Center for PTSD, VA Boston Healthcare System, Boston, MA, USA; fDepartment of Psychiatry, Boston University School of Medicine, Boston, MA, USA; gDepartment of Psychology, University of Arizona, Tucson, AZ, USA

ABSTRACT

Background: This study was an examination of the puzzling finding that people assessed for symptoms of posttraumatic stress disorder (PTSD) consistently score higher on the self-report PTSD Checklist for *DSM-5* (PCL-5) than the Clinician-Administered PTSD Scale for *DSM-5* (CAPS-5). Both scales purportedly assess PTSD severity with the same number of items, scaling, and scoring range, but differences in scores between measures make outcomes difficult to decipher.
Objective: The purpose of this study was to examine several possible psychometric reasons for the discrepancy in scores between interview and self-report.
Method: Data were combined from four clinical trials to examine the baseline and posttreatment assessments of treatment-seeking active duty military personnel and veterans.
Results: As in previous studies, total scores were higher on the PCL-5 compared to the CAPS-5 at baseline and posttreatment. At baseline, PCL-5 scores were higher on all 20 items, with small to large differences in effect size. At posttreatment, only three items were not significantly different. Distributions of item responses and wording of scale anchors and items were examined as possible explanations of the difference between measures. Participants were more likely to use the full range of responses on the PCL-5 compared to interviewers.
Conclusions: Suggestions for improving the congruence between these two scales are discussed. Administration of interviews by trained assessors can be resource intensive, so it is important that those assessing PTSD severity are afforded confidence in the equivalence of their assessment of PTSD regardless of the assessment method used.

Una comparación entre el CAPS-5 y PCL-5 para evaluar el TEPT en muestras de militares y veteranos en búsqueda de tratamiento

Antecedentes: Este estudio fue una examinación del hallazgo desconcertante de que las personas evaluadas por síntomas de trastorno de estrés postraumático (TEPT) obtienen consistentemente puntuaciones más altas en la Lista de verificación de TEPT de auto reporte según el DSM-5 (PCL-5 en su sigla en inglés) que en la Escala de TEPT administrada por un médico según el DSM-5 (CAPS-5 en su sigla en inglés). Ambas escalas supuestamente evalúan la gravedad del TEPT con la misma cantidad de ítems, escalas y rango de puntaje, pero las diferencias en los puntajes entre las medidas hacen que los resultados sean difíciles de descifrar.
Objetivo: El propósito de este estudio fue examinar varias posibles razones psicométricas de la discrepancia en las puntuaciones entre la entrevista y el auto reporte.
Método: Los datos fueron de la combinación de de cuatro ensayos clínicos para examinar las evaluaciones iniciales y posteriores al tratamiento del personal militar en servicio activo y los veteranos en búsqueda de tratamiento.
Resultados: Al igual que en estudios anteriores, las puntuaciones totales fueron más altas en el PCL-5 en comparación con el CAPS-5 al inicio y después del tratamiento. Al inicio del estudio, las puntuaciones de PCL-5 fueron más altas en los 20 ítems, con diferencias pequeñas a grandes en el tamaño del efecto. En el postratamiento, solo tres ítems no fueron significativamente diferentes. Se examinaron las distribuciones de las respuestas a los ítems y la redacción de las escalas y los ítems como posibles explicaciones de la diferencia entre las medidas. Los participantes tenían más probabilidades de utilizar la gama completa de respuestas en el PCL-5 en comparación con los entrevistadores.
Conclusiones: Se discuten sugerencias para mejorar la congruencia entre estas dos escalas. La administración de entrevistas por evaluadores capacitados puede requerir muchos recursos,

HIGHLIGHTS

- The purpose of this study was to examine two commonly used measures of posttraumatic stress disorder, the Clinician-Administered PTSD Scale (CAPS-5), an interview measure, and the PTSD Checklist (PCL-5), a self-report measure, to explore discrepancies in scores.
- Both measures have the same number of items and range of scores assessing the identical 20 symptoms of PTSD, yet higher scores are reported on the PCL-5.
- It appears that the differences in wording of the anchors may contribute to discrepancies in scoring.
- Addressing these problems would allow for a better match in scoring between scales.

CONTACT Patricia A. Resick ✉ patricia.resick@duke.edu 🏢 Department of Psychiatry and Behavioral Sciences, Duke Health, Durham, NC, USA

por lo que es importante que aquellos que evalúan la gravedad del TEPT tengan confianza en la equivalencia de su evaluación del TEPT, independientemente del método de evaluación utilizado.

比较 CAPS-5 和 PCL-5 评估军人和退伍军人寻求治疗样本中的 PTSD

背景：本研究检验了一个令人费解的发现，即评估创伤后应激障碍 (PTSD) 症状的人在 DSM-5 自我报告 PTSD 检查表 (PCL-5) 中的得分始终高于 DSM-5临床医生管理 PTSD 量表 – 5（CAPS-5）。据称，这两种量表都使用相同数量的条目、量表和评分范围来评估 PTSD 严重程度，但测量之间的分数差异使得结果难以解读。
目的：本研究旨在考查访谈和自我报告之间分数差异的几个可能的心理测量学原因。
方法：数据是四项临床试验的综合数据，用于考查寻求治疗的现役军人和退伍军人的基线和治疗后评估。
结果：与前人研究一样，在基线时和治疗后，PCL-5 的总分相较于 CAPS-5 更高。 在基线时，所有 20个条目的 PCL-5 分数都较高，差异的效应量大小不等。 在治疗后，只有三个条目没有显著差异。 考查了条目响应的分布和量表锚点和条目的措辞，作为对措施之间差异的可能解释。与访谈者相比，参与者更有可能在 PCL-5 上使用全范围的回答。
结论：讨论了改进这两个量表之间一致性的建议。由训练有素的评估员进行面谈可能会占用大量资源，因此重要的是，无论使用何种评估方法，评估 PTSD 严重程度的方法都必须具备对 PTSD 评估的等效性。

1. Introduction

When assessing posttraumatic stress disorder (PTSD) for research or clinical work, there are both structured clinician interviews and self-report measures (e.g. Weathers et al., 2013a; Weathers et al. 2013). Typically for research, the clinician-administered interview is considered the 'gold standard' (e.g. Weathers et al., 2001). Although it still relies on self-report, interviewers can ask follow-up questions to clarify responses and use clinical judgement when scoring severity. However, in clinical practice, it is usually not feasible to administer clinical interviews, which might take an entire 60-minute session to administer, especially when a limited number of sessions are allowed. Interviews may also be impractical in large studies due to limited resources or concern about assessment burden. Self-report measures are quick to administer and are frequently given throughout treatment to monitor changes in symptoms. Because self-report measures are often used to assess PTSD in both research and clinical practice because of their practicality, it is important that there is an understanding of how results of self-report measures compare to clinical interviews (Bovin & Weathers, 2022; Livingston et al., 2021).

The Clinician-Administered PTSD Scale-5 (CAPS-5; Weathers et al., 2013a) and the PTSD Checklist (PCL-5; Weathers et al. 2013) are gold-standard measures for the assessment of PTSD. However, correlations between these interview and self-report measures have varied widely even though they are purportedly measuring the same construct. The original versions of the CAPS (Blake et al., 1990) and PCL (Weathers et al., 1993) were based on the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*; American Psychiatric Association [APA], 1994). Correlations between total scores on the CAPS and PCL have ranged from .30 to .93 (Adkins et al., 2008; Blanchard et al., 1996; Bollinger et al., 2008; Forbes et al., 2001; Keen et al., 2008). Forbes et al. (2001) found that item-level correlations ranged from .07 to .57. At a 9-month treatment follow-up, the total score correlation improved to .62, but the item-level correlations were .06 to .57. Thus, there was not consistent agreement between clinician assessment and self-report of the same symptoms. In these past versions of the measures, the interview and self-report measures of PTSD differed in the numbers of items, scoring strategies, and wording. For example, total scores for CAPS interview ranged from 0 to 136, while the self-report PCL ranged from 17-85. This meant that scores on these measures could not be directly compared.

With revisions of both CAPS (CAPS-5; Weathers et al., 2018) and PCL (PCL-5; Blevins et al., 2015) for *DSM-5* (APA, 2013), an effort was made to align the scoring of the measures. In *DSM-5,* PTSD has 20 symptoms divided into four clusters. The CAPS-5 stopped separating frequency and intensity scoring and moved to a single severity score to match the PCL-5. In the *DSM-5* version, PCL-5 item scoring was adjusted to start at 0 to match scoring for the CAPS-5. Both scales now have 20 items that are each scored 0–4, totalling 0–80. With these revisions, the PCL-5 and CAPS-5 should be measuring the same construct (severity of PTSD symptoms) on the same scale; therefore, item and total scores should show stronger alignment than did the *DSM-IV* versions. However, in studies that have included both measures, the PCL-5 typically has resulted in scores 8–20 points higher at baseline or nontreatment single assessment. For example, examining the psychometric properties of the PCL-5 with predominantly male veterans, Bovin et al. (2016) found PCL-5 scores were on

average 8.2 points higher than the CAPS-5. Lee et al. (2022) found the PCL-5 to be 8.8 points higher with male veterans. Krüger-Gottschalk et al. (2022) found that, across five German treatment centres with mixed genders, there was an average of 10.3 points higher total scores on the PCL-5 compared to the CAPS-5.

Researchers have also reported scores on the CAPS-5 and PCL-5 at baseline and posttreatment, offering opportunities to examine the relationship between the measures across treatments. For example, Davis et al. (2020) compared two treatments in a sample of veteran (92%) and civilian adults with PTSD. At pretreatment, the PCL-5 total scores were 8.1 and 9.1 points higher than the CAPS-5 in each treatment group, and at posttreatment, the PCL-5 scores were 7.2 and 9.2 points higher. In a study comparing dialectical behaviour therapy for PTSD (DBT-PTSD) to CPT for civilians in Germany, Bohus et al. (2020) found that, at pretreatment, the PCL-5 scores were on average 9.5 and 8.6 points higher than the CAPS-5 for DBT-PTSD and CPT, respectively. At posttreatment, the PCL-5 scores were 3.3 and 7.3 points higher for the DBT-PTSD and CPT groups, respectively. Overall, PCL-5 total scores were consistently higher than CAPS-5 total, so discrepancies in scores are likely not attributable to specific treatment effects.

Other studies with CAPS-5 and PCL-5 scores at pre- and posttreatment suggest that total scores for these measures may converge after treatment exposure. De Jongh et al. (2020) reported on civilian patients in the Netherlands (82% female) who received treatment for PTSD and borderline personality disorder with prolonged exposure and eye movement desensitization and reprocessing. Average pretreatment scores on the PCL-5 were 7.7 points higher than on the CAPS-5 and at posttreatment were 3.4 points higher. Held et al. (2022) found with a community sample of predominantly women (71%) a 15.8 higher mean score on the PCL-5 compared to the CAPS-5 at pretreatment but a 2.8 point difference at posttreatment. Following treatment the measures were in closer agreement probably due to reduction in symptoms.

Overall, we could find no studies in which the PCL-5 did not have higher mean scores than the CAPS-5. One could argue that the patients are overreporting their symptoms on the self-report measure, but symptom exaggeration would likely inflate scores on the CAPS-5 interview as well. Also, counter to expectations that patients might exaggerate symptoms, Monson et al. (2008) found that patients reported greater decreases in symptoms during treatment than did interviewers. That rationale also would not explain the discrepancy found in the psychometric studies without treatment-seekers. Finally, the closer match between assessments at posttreatment further erodes hypotheses about symptom overreporting specific to self-report. As patients recover from their PTSD, their range of scores decreases regardless of the type of assessment.

Lower scores on the CAPS-5 may be attributable to differences in clinician training for CAPS-5 administration, but the studies cited found similar results despite being from different research groups. One possibility that has not been examined is that there may be problems with the measures themselves, such that the self-report scale pulls for higher scores and the interview for lower scores. If this is the case, then some people (i.e. those seeking treatment in clinical trials with inclusion criteria based on the CAPS-5) may be refused treatment, or the studies and clinics that use only the self-report measure may not accurately represent PTSD.

The purpose of this paper was to examine data from four samples of treatment-seeking active duty service members and veterans collected by different researchers within the same Network of affiliated studies (Peterson et al., 2021) including item-level and total congruence between the CAPS-5 and PCL-5. The question of discrepancies between interview and self-report assessments may be particularly relevant in a military sample. For instance, military values such as stoicism may result in underreporting of symptoms, while a desire to establish disability status may lead to overreporting. All of the analyses were exploratory and focused on examining potential sources of discrepancy, so no specific hypotheses were made.

2. Method

2.1. Participants and procedures

The sample for this secondary analysis consisted of 739 treatment-seeking, active duty military members or veterans from four studies who were assessed with the CAPS-5 and PCL-5 to determine if they met study inclusion criteria. Unlike most treatment studies comparing the CAPS and PCL, we included everyone who completed both the baseline CAPS-5 and PCL-5 regardless of whether they were enrolled into the study, resulting in a wider range of scores reflecting syndromal and subsyndromal PTSD. Posttreatment data from the intent-to-treat samples were examined (i.e. including all randomized patients within the trials). If someone dropped out of treatment prematurely, they were asked to return for assessment at the time they would have otherwise been assessed. All participants assessed were included in the current analyses to increase the range of responses. The posttreatment comparison of CAPS-5 and PCL-5 was composed of participants who were enrolled and treated in one of the studies and completed a posttreatment assessment with both measures. This posttreatment sample included 266

Table 1. Combined baseline and posttreatment data from included studies.

		Baseline					Posttreatment			
		CAPS-5		PCL-5			CAPS-5		PCL-5	
Study	Total *N*	PTSD *N*	PTSD %	PTSD *N*	PTSD %	Total *N*	PTSD *N*	PTSD %	PTSD *N*	PTSD %
McGeary et al. (2022)	225	175	77.8	180	80.0	92	58	63.0	57	62.0
Peterson et al. (2022)	154	123	79.9	123	79.9	61	30	49.2	26	42.6
Resick et al. (2021)	185	149	80.5	152	82.2	83	36	43.4	33	39.8
Taylor et al. (2022)	175	139	79.4	128	73.1	30	23	76.7	20	66.7
Total	**739**	**586**	**79.3**	**589**	**78.9**	**266**	**147**	**55.3**	**136**	**51.1**

Notes. Individuals with any missing data on the Clinician-Administered PTSD Scale for *DSM-5* (CAPS-5) or PTSD Checklist for *DSM-5* (PCL-5) were excluded from analyses. PTSD = posttraumatic stress disorder.

participants (see Table 1). The posttreatment sample varied in treatment response and therefore reflected both PTSD and non-PTSD diagnostic status.

The majority of participants were married (69.4%), White (51.4%), and male (85.3%). Most participants were in the U.S. Army (84.0%) with an average of 13.88 (*SD* = 7.36) years of service. Approximately half (50.8%) of the sample reported being active duty at the time of the baseline assessment, 46.9% were veterans, and the rest were National Guard/Reservists. Almost all (98.2%) had at least one prior deployment. Tables 2 and 3 list detailed information on sociodemographics, military and trauma characteristics of the sample. Although all of the studies were preregistered with clinicaltrials.gov for the clinical trials, this secondary study was not preregistered.

2.2. Measures and assessment procedure

2.2.1. Training on administering the CAPS-5
The CAPS-5 was administered by master's- or doctoral-level staff who received standardized training from one of the primary authors of the CAPS-5 to a criterion of reliability; all subsequently received certification as independent evaluators following a minimum of two assessments recorded and evaluated with re-training if indicated until they achieve 80% reliability with the master evaluator (Barnes et al., 2019; Peterson et al., 2021). Evaluators were masked to treatment status at all assessments. As part of studies' training and fidelity assessment, approximately 3% of CAPS-5 interviews were randomly selected for review on twice monthly calibration calls to establish interrater reliability and prevent drift. The interrater reliability of CAPS-5 case decisions was excellent (*N* = 78; Cohen's kappa = 0.90) as was the correlation of severity scores between raters (*r* = 0.98).

2.2.2. Selection of index events for the CAPS-5
Prior to completing the CAPS-5, assessors identified the Criterion A event on which to anchor the participants' responses. This interview was modified to more thoroughly assess trauma experiences, including those in a military context. The Selection of Index Event interview followed self-report measures regarding combat and postbattle experiences using the Deployment Risk and Resiliency Inventory-2 (DRRI-2)

Table 2. Baseline sociodemographic and military characteristics of combined samples.

Variable	N	%
Gender		
Male	622	85.3
Female	107	14.7
Race		
Hispanic/Latinx	225	30.7
White, Non-Hispanic	216	29.5
Black	193	26.4
Biracial	35	4.8
Asian/Pacific Islander	27	3.7
Native American	16	12.2
Other[a]	20	2.7
Marital status		
Married	506	69.4
Divorced/separated	116	15.9
In a relationship	66	9.0
Single	41	5.6
Education		
Some HS to HS equivalent[b]	97	13.3
Some college to associates degree	735	60.1
4-year college degree	142	19.5
Graduate degree	52	7.1
Branch		
Army	613	84.0
Marines	45	6.2
Air Force	39	5.3
Navy	32	4.4
Coast Guard	1	< 1.0
Military status		
Active Duty	371	50.8
Veteran	342	46.9
Reserve	17	2.3
Rank		
E1 to E3	13	1.8
E4 to E6	468	64.5
E7 to E9	181	2.5
WO1 to O6	64	8.8
Duty type		
Combat Arms	298	41.3
Combat Support	154	21.3
Combat Service Support	270	37.4
# of Deployments		
0	13	1.8
1	201	28.0
2	194	27.1
3+	309	43.1
	M	*SD*
Age	38.52	8.44
Years of Service	13.88	7.36

Notes: Cell counts and proportions vary based on available data across variables. HS = High School; E-1 to E-3 = junior enlisted officer; E-4 to E-6 = junior noncommissioned officer; E-7 to E-9 = senior noncommissioned officer; O = officer; WO = warant officer.
[a]Other Race category was comprised of individuals who identified with a nationality (e.g. 'American'; 'European') or selected 'Other' and did not provide a response.
[b]HS equivalent includes HS degree or General Education Development test.

subscales (Vogt et al., 2013) and the Life Events Checklist for *DSM-5* (Weathers et al., 2013b). The

interview ensured that the most distressing event was chosen among the experiences that met PTSD Criterion A. The same event was used to anchor the responses on the PCL-5.

2.2.3. Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)

The CAPS-5 (Weathers et al., 2013a; Weathers et al., 2018) is a clinical diagnostic interview with items reflecting *DSM-5* PTSD symptoms rated on a 5-point scale to assess PTSD diagnosis and symptom severity: 0 (*absent*), 1 (*mild/subthreshold*), 2 (*moderate/threshold*), 3 (*severe/markedly elevated*) and 4 (*extreme/incapacitating*). The CAPS-5 has demonstrated internal consistency ($\alpha = 0.88$), interrater reliability (kappa = 0.78), and test-retest reliability (kappa = 0.83; Weathers et al., 2018). All participants met the duration and either distress or impairment questions in the CAPS-5 so these items were not included in diagnosis for this purpose. They are included in Table 3 for descriptive purposes.

2.2.4. PTSD Checklist for DSM-5 (PCL-5)

The PCL-5 (Blevins et al., 2015; Weathers et al., 2013) is a 20-item self-report measure that assesses *DSM-5* PTSD symptoms using a severity rating Likert-type scale: 0 (*not at all*), 1 (*a little bit*), 2 (*moderate*), 3 (*quite a bit*) and 4 (*extremely*). The PCL-5 assesses how much participants have been bothered by PTSD symptoms related to a specific event in the past month or past week. The PCL-5 has demonstrated good internal consistency ($\alpha = 0.96$) and test-retest reliability (0.84; Bovin et al., 2016).

2.2.5. Assessment standard operating procedure

All four studies were conducted under the auspices of the STRONG STAR Consortium and the Consortium to Alleviate PTSD (Peterson et al., 2021). Studies were conducted in Killeen, Texas, at Fort Cavazos (formerly named Fort Hood) and in San Antonio, Texas. All studies used the same assessment

Table 3. Time since trauma, distress/impairment, and trauma type sample descriptives.

Descriptive trauma information	M	SD
Years since trauma	10.00	6.43
CAPS 5 Total Distress/Impairment	6.58	2.14
Subjective Distress	2.37	0.60
Social Functioning Impairment	2.39	0.98
Occupational Functioning Impairment	1.82	1.04
Types of trauma	N	%
Life Threat to Self	221	29.9
Life Threat to Others	112	15.2
Aftermath of Violence	152	20.6
Traumatic Loss	182	24.6
Moral Injury by Self	36	4.9
Moral Injury by Others	36	4.9

Notes. Types of Trauma are based on Stein et al. (2012) Trauma Type Taxonomy.

procedures and trainings. The CAPS-5 followed the selection of index event determining the referent event. The PCL-5 was completed following other self-report measures administered after the CAPS-5 in order to separate participants' responses from the CAPS-5. Participants were prompted to consider the previously identified index event when completing the PCL-5, and for all assessments during and posttreatment.

2.3. Data analysis plan

Analyses included data from anyone who completed the PCL-5 and CAPS-5 at the baseline assessment. Participants with missing data on either measure were excluded from analyses because one of our aims was to investigate item-level differences between the PCL-5 and CAPS-5. Prior to listwise deletion (complete case analysis), we employed Little's (1998) missing completely at random (MCAR) test, which revealed missingness was MCAR. All individuals were seeking treatment for PTSD or a trauma-related comorbidity at baseline, which can lead to restriction of range. Therefore, all analyses were replicated with the subsample of individuals who also completed the PCL-5 and CAPS-5 at the posttreatment assessment and had no missing data on either measure.

We first examined diagnostic agreement (Cohen's κ) between the PCL-5 and CAPS-5. For both measures, a PTSD symptom was determined to be present if the corresponding item had a rating of 2 or higher, reflecting 'moderate' or 'threshold' severity, and, consistent with *DSM-5* criteria, a positive diagnosis was indicated by the presence of at least one intrusion symptom, one avoidance symptom, two cognition and mood symptoms, and two hyperarousal symptoms. A Cohen's κ below .60 can be considered weak agreement, greater than .60 suggests moderate agreement, .80 to .90 indicates strong agreement, and scores greater than .90 suggest near perfect agreement (McHugh, 2012). Internal consistency (Cronbach's α) was also calculated to determine the interrelatedness of measure items on the PCL-5 and CAPS-5. Cronbach's α greater than .70 is generally considered acceptable, with higher scores suggestive of better internal consistency (Tavakol & Dennick, 2011).

Next, we evaluated the convergent validity of the PCL-5 and CAPS-5 total and item-level scores using the Pearson's *r* correlation coefficient. A correlation coefficient greater than .50 can reflect good convergent validity, but estimates greater than .70 are typically recommended as strong evidence of convergent validity when both instruments are hypothesized to measure the same construct (Allen & Yen, 2002). The correction for attenuation (δ) method ($\delta_{xy} = r_{xy}/$

sqrt $[\alpha_x * \alpha_y]$) was calculated to estimate the relationship between the instrument total scores assuming the absence of measurement error. A series of paired sample t tests was used to examine mean differences in the total and item scores at baseline and posttreatment on the PCL-5 and CAPS-5 and to evaluate change score differences between instruments following treatment. Cohen's d was computed to determine the magnitude of differences between measures. Finally, frequency (%) statistics were calculated to evaluate the distribution of Likert ratings across items on both measures. All analyses were completed using SPSS Version 27.0. The data from this study are maintained at the University of Texas Health Science Center at San Antonio in the STRONG STAR Repository. Requests for access to the data as well as for materials and the analysis code can be emailed to repository@strongstar.org.

3. Results

3.1. Diagnostic agreement

Most of the sample met criteria for PTSD on the CAPS-5 (79.3%) and the PCL-5 (79.9%) at baseline (Table 1). At baseline, diagnostic reliability was below the recommended cut-off for good agreement ($\kappa = .55$), with diagnostic agreement occurring 84.9% of the time between the CAPS-5 and PCL-5. Among those randomized to a study, 44.7% achieved PTSD remission on the CAPS-5 and 48.9% no longer had probable PTSD on the PCL-5 following treatment. At posttreatment, there was moderate agreement ($\kappa = .74$), with a diagnostic consistency rate of 86.8%.

3.2. Internal consistency and convergent validity

Baseline alpha coefficients for internal consistency (α) were .83 and .92 on the CAPS-5 and PCL-5, respectively, both in the acceptable range. As shown in Table 4, the correlation between the total scores obtained on both measures was acceptable, r (739) = .75, $\delta = .85$, $p < .001$, and the correlations between the item (symptom) level scores obtained on both measures ranged from $r = .46$ to .66 (all ps < .001). Internal consistency at posttreatment increased to $\alpha = .91$ on the CAPS-5 and $\alpha = .97$ on the PCL-5, both in the excellent range (Table 4). The correlation between the total scores obtained on both measures increased to good, r (266) = .86, $\delta = .91$, and the correlation between the item-level scores obtained on both measures also increased and ranged from acceptable to good (i.e. .49 to .76; all ps < .001).

3.3. Total, item, and change score differences

All baseline scores were higher on the PCL-5 compared to the CAPS-5 (all p's < .001), with a total score mean difference of 13.95, $p < .001$, $d = 1.35$ (Table 4). With the exception of items 9 (negative beliefs), 11 (negative emotions), and 20 (sleep), the item and total scores remained significantly higher on the PCL-5 compared to the CAPS-5 following treatment (see Table 5). Differences between the PCL-5 and CAPS-5 at posttreatment (d ranged from < −0.01 to 0.76) were smaller than baseline observations (d ranged from 0.26 to 1.03). Regarding instrument differences in change, participants had

Table 4. Baseline CAPS-5 and PCL-5 Item and Total Score Differences and Correlations.

	PCL-5		CAPS-5					
	M	SD	M	SD	M_{diff}	T	D	r
Total Score	46.54	15.46	32.59	10.57	13.95	36.69**	1.35	0.75**
1. Unwanted memories	2.49	1.01	2.01	1.00	0.48	12.83**	0.47	0.50**
2. Dreams	2.19	1.23	1.75	1.21	0.44	11.84**	0.44	0.66**
3. Flashbacks	1.66	1.23	0.48	0.85	1.18	26.91**	0.99	0.39**
4. Distress at reminders	2.59	1.12	1.75	1.00	0.85	21.09**	0.78	0.47**
5. Cued physical reactions	2.38	1.21	1.19	1.12	1.19	24.71**	0.91	0.37**
6. Avoid memories	2.65	1.18	1.82	1.03	0.83	19.49**	0.72	0.46**
7. Avoid external	2.66	1.24	1.86	1.19	0.80	17.88**	0.66	0.50**
8. Amnesia	1.34	1.40	0.37	0.83	0.98	22.09**	0.81	0.52**
9. Negative beliefs	2.08	1.38	1.73	1.38	0.36	7.21**	0.27	0.53**
10. Blame self/others	1.86	1.49	1.07	1.31	0.79	16.97**	0.62	0.60**
11. Negative emotions	2.47	1.20	2.12	1.01	0.35	8.62**	0.32	0.51**
12. Lost interests	2.73	1.21	2.04	1.21	0.69	16.35**	0.60	0.55**
13. Detachment	2.76	1.21	2.34	1.20	0.41	10.48**	0.39	0.61**
14. No positive emotions	2.25	1.32	1.74	1.33	0.51	12.83**	0.47	0.67**
15. Aggression	2.45	1.20	1.68	0.94	0.77	18.98**	0.70	0.48**
16. Self-destructive	0.75	1.11	0.30	0.76	0.46	13.20**	0.49	0.55**
17. Hypervigilance	2.90	1.14	2.28	0.96	0.62	15.22**	0.56	0.46**
18. Startle	2.37	1.31	1.17	1.03	1.19	28.02**	1.03	0.53**
19. Concentration	2.78	1.18	2.05	1.10	0.73	19.74**	0.73	0.61**
20. Sleep	3.17	1.03	2.83	0.99	0.35	9.70**	0.36	0.53**

Notes. $N = 739$. PCL-5 = PTSD Checklist for DSM-5; CAPS-5 = Clinician-Administered PTSD Scale for DSM-5; r = PCL-5 and CAPS-5 correlation; total score correction for attenuation (δ) correlation = .85; M_{diff} = PCL-5 and CAPS-5 Mean difference; d = Cohen's d effect size.
**$p < .001$. The prompts at the left are not the literal items from the DSM-5 but abbreviations of the concepts.

Table 5. Posttreatment CAPS-5 and PCL-5 Item and Total Score Differences and Correlations.

	PCL-5		CAPS-5					
	M	*SD*	*M*	*SD*	*M*$_{diff}$	*t*	*D*	*r*
Total Score	33.65	20.89	26.29	13.96	7.36	10.44**	0.64	.86**
1. Unwanted memories	1.76	1.22	1.52	1.13	0.24	4.38**	0.27	.71**
2. Dreams	1.6	1.31	1.38	1.23	0.22	4.13**	0.25	.76**
3. Flashbacks	1.23	1.25	0.4	0.8	0.83	12.46**	0.76	.51**
4. Distress at reminders	1.85	1.32	1.4	1.09	0.45	7.21**	0.44	.65**
5. Cued physical reactions	1.74	1.34	1.14	1.15	0.6	8.64**	0.53	.59**
6. Avoid memories	1.88	1.42	1.36	1.16	0.52	7.55**	0.46	.64**
7. Avoid external	1.94	1.47	1.48	1.22	0.46	7.10**	0.44	.71**
8. Amnesia	0.87	1.19	0.3	0.81	0.57	9.25**	0.57	.55**
9. Negative beliefs	1.41	1.43	1.36	1.33	0.05	0.8	0.05	.70**
10. Blame self/others	1.18	1.37	0.81	1.15	0.37	5.07**	0.31	.57**
11. Negative emotions	1.58	1.33	1.58	1.13	< 0.01	−0.06	< − 0.01	.70**
12. Lost interests	2.04	1.42	1.55	1.33	0.49	7.45**	0.46	.69**
13. Detachment	1.97	1.43	1.8	1.37	0.17	2.71*	0.17	.73**
14. No positive emotions	1.71	1.4	1.39	1.3	0.32	5.13**	0.31	.72**
15. Aggression	1.83	1.3	1.47	0.98	0.36	5.62**	0.34	.62**
16. Self-destructive	0.58	1	0.25	0.7	0.33	5.96**	0.37	.49**
17. Hypervigilance	2.07	1.42	1.87	1.2	0.2	2.93*	0.18	.64**
18. Startle	1.72	1.39	0.97	1.06	0.75	10.47**	0.64	.57**
19. Concentration	2.11	1.41	1.79	1.23	0.32	5.20**	0.32	.72**
20. Sleep	2.6	1.35	2.49	1.24	0.11	1.62	0.1	.62**

Notes. $N = 266$. PCL-5 = PTSD Checklist for *DSM-5*; CAPS-5 = Clinician-Administered PTSD Scale for *DSM-5*; r = PCL-5 and CAPS-5 correlation; total score correction for attenuation (δ) correlation = .91; M_{diff} = PCL-5 and CAPS-5 Mean difference; d = Cohen's d effect size.
*$p < .01$; **$p < .001$.

greater reductions on the PCL-5 ($M = 15.53$, SD = 18.22) compared to the CAPS-5 ($M = 8.92$, SD = 12.41) following treatment ($M_{diff} = 6.61$, $p < .001$, $d = 0.57$).

3.4. CAPS-5 and PCL-5 item-level distributions

Regarding item-level response distribution, PTSD symptoms were most commonly rated as 0 (*absent*; 32.0%) or 2 (*moderate/threshold*; 34.2%) on the CAPS-5, whereas 3 (*quite a bit*; 29.7%) or 4 (*extremely*; 22.6%) were the most common ratings on the PCL-5 (Figure 1). Additionally, the rating distribution was bimodal on the CAPS-5, while the response distribution was more equally represented on the PCL-5. That is, ratings of 1 (*mild*; 5.4%) and 4 (*extreme/*

incapacitating; 4.0%) were infrequently coded on the CAPS-5, whereas all five response anchors, which are administered with slightly different wording, were more equally used on the PCL-5 and ranged between 14.1% to 29.7%.

Item-level response distribution trends at post-treatment demonstrated an overall shift toward lower ratings compared to baseline observations on both measures (Figure 2). The most common PTSD symptom ratings on the CAPS-5 remained 0 (*absent*; 42.6%) or 2 (*moderate/threshold*; 30.0%), whereas the most common responses were 0 (*not at all*; 28.3%) or 1 (*a little bit*; 21.5%) on the PCL-5 (Figure 2). Similar to baseline, the item rating distribution was largely bimodal on the CAPS-5, and the ratings 1 (*mild*; 6.5%) and 4 (*extreme/*
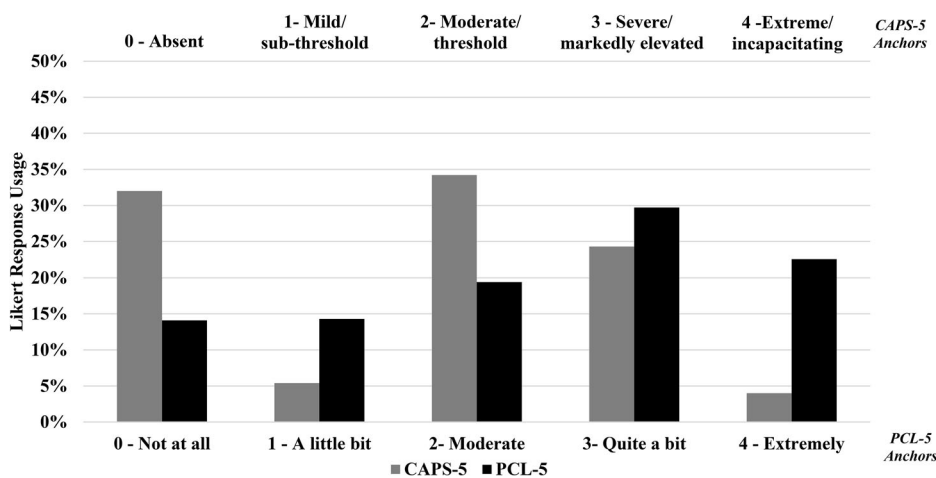

Figure 1. Baseline PCL-5 and CAPS-5 Likert Scale Rating Proportions Across All Items. *Notes*. PCL-5 = PTSD Checklist for DSM-5; CAPS = Clinician-Administered PTSD Scale for DSM-5. CAPS-5 anchor descriptions are listed at the top of the figure; PCL-5 anchor descriptions are listed at the bottom of the figure. Percent (%) represents the frequency of selecting Likert rating across all 20 items of the PCL-5 and CAPS-5, respectively.
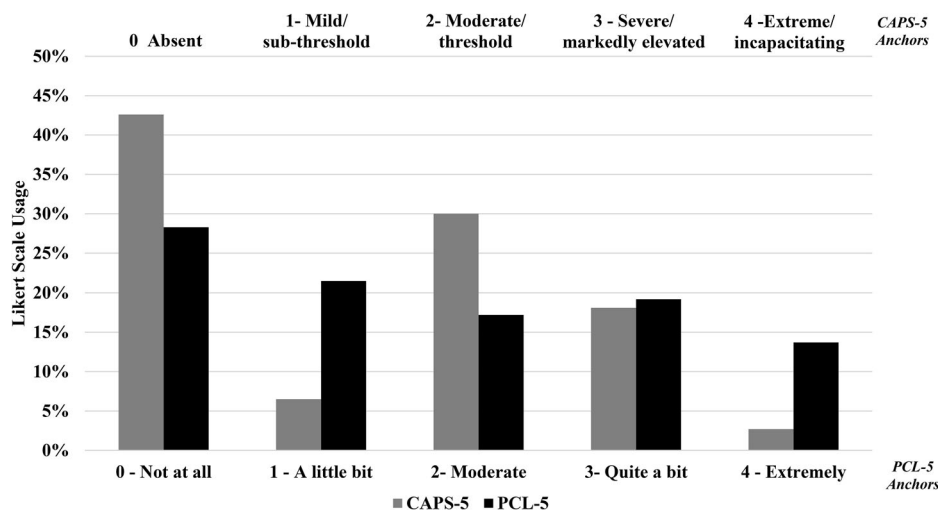
Figure 2. Posttreatment PCI-5 and CAPS-5 Likert Scale Rating Proportions Across All Items. *Notes.* PCL-5 = PTSD Checklist for DSM-5; CAPS-5 = Clinician-Administered PTSD Scale for DSM-5. CAPS-5 anchor descriptions are listed at the top of the figure; PCL-5 anchor descriptions are listed at the bottom of the figure. Percent (%) represents the frequency of selecting Likert rating across all 20 items of the PCL-5 and CAPS-5, respectively.

incapacitating; 2.7%) were infrequently used. Again, the five anchors were more equally represented on the PCL-5 (13.7% to 28.3%).

4. Discussion

This study examined item-level and total score differences between the CAPS-5 (Weathers et al., 2013a) and PCL-5 (Weathers et al., 2013) in a large sample of active duty military and veterans seeking treatment across four clinical trials. Although these measures were designed to assess the same construct (i.e. PTSD symptom severity), and are now on the same 0–4 scale, the results showed that items on the self-report PCL-5 were consistently rated higher than the corresponding items on the clinician-administered CAPS-5.

Across the four treatment-seeking samples at baseline, participants scored an average of 14 points higher on the PCL-5 than the interviewers scored them on the CAPS-5, and scores were significantly different on every individual item, with effect sizes ranging from 0.27 to 1.03. However, correlations within and between measures were within the acceptable range at baseline.

Consistent with prior research, these differences were less striking at posttreatment. This may be attributed to the smaller sample, although $n = 266$ is larger than most studies, and there was still a 7-point difference in scores. At posttreatment, patients had improved on the most severe scores in their self-reports. The two measures converged on only three items at posttreatment: negative beliefs, negative emotions, and sleep. What was clear at both time points was that interviewers preferentially scored 0s and 2s on the CAPS-5, whereas participants seem to have used a more equal distribution of responses

(favouring elevated symptoms at baseline and favouring remission at post).

Because other studies from a range of locations have found the same phenomenon of higher PCL-5 scores (e.g. Bohus et al., 2020; Bovin et al., 2016; Davis et al., 2020; Krüger-Gottschalk et al., 2022; Lee et al., 2022), poor interviewer training unique to any one study might be ruled out. In addition, because of the decrease in symptoms at posttreatment and the reliance on self-report for both measures, it is probably not just an exaggeration of distress on the part of participants. One possible explanation is the scaling of the items. Although both scales have a 5-point scale, for the PCL-5, the anchor for the '4' rating is labelled *extremely* while on the CAPS-5, the anchor for the '4' rating is labelled *extreme/incapacitating* (Weathers et al., 2013a; Weathers et al., 2013). CAPS-5 interviewers were trained to reserve the use of *extreme/incapacitating* to only extreme and/or incapacitating symptom reports. No such instructions are provided on the self-report PCL-5.

The wording '*extreme/incapacitating*' may seem to not apply to active duty military members. An interviewer might reason that if their symptoms were incapacitating, they would not be able to work. The word 'incapacitated' may be interpreted as more severe than 'extremely,' which is used on the PCL-5, and which may explain why the same participants in the study were willing to endorse the maximum score for more than 20% of the items on the self-report form. Furthermore, the CAPS-5 only gives more specific item scoring guidance on two of the scale anchors, 'moderate' and 'severe,' but does not provide guidance on the specific item-level criteria for subthreshold or extreme/incapacitated scores (Weathers et al., 2013a). The self-report wording for scores of 4 (*extremely*) may better match the CAPS-5 wording for

scores of *3 (severe)*. Perhaps more equivalent scaling and anchors on the two forms is in order.

Aside from the nonequivalence of the rating scales, the different wording of items on the CAPS-5 and PCL-5 could have meaningful differences. For example, on the item with the largest effect-size difference at baseline in this study, psychogenic amnesia, the CAPS-5 asks follow-up questions and rules out aetiology due to head injury or intoxication (Weathers et al., 2013a). The PCL-5 merely asks about 'Trouble remembering important parts of the stressful experience' without adding 'not due to head injury or substance use' (Weathers et al., 2013). This wording could be added for greater clarity and consistency. Amnesia due to head injury or substance ingestion cannot be changed with treatment because the event is not stored in long-term memory, whereas dissociative amnesia can be reduced with treatment.

Another item on the PCL-5 that seems less problematic in this military sample but has previously been a source of confusion is 'Blaming yourself or someone else for the stressful experience or what happened after it,' which is ostensibly intended to assess distorted cognitions about self or other blame (Weathers et al., 2013). On the CAPS-5 there are a series of questions about blame of self and others and their appraisal of whether self/other caused the event. In samples with clear individual perpetrators, such as rape, child maltreatment, or intimate-partner violence, the lack of clarity of the PCL-5 item could be problematic. For example, at pretreatment, victims of rape may blame themselves and score themselves at 4 (*extremely*); but at posttreatment, they could switch and rightfully blame the perpetrator and also score it 4. The scoring would indicate no change in symptom severity despite a good response to treatment. An easy solution would be to ask about 'Blaming yourself or someone else (who didn't intend the harm) for the stressful experience or what happened after it.'

Another possible source of the scoring discrepancies may be that participants had difficulties attributing their symptoms to a specific trauma when completing the self-report measures. Unlike the CAPS-5 assessment, in which the interviewer can clarify the source of the reported symptoms, participants may have responded to the PCL with regard to symptoms resulting from other comorbid diagnoses, rather than tied to the index traumatic event, resulting in higher levels of more generalized symptomatology on the self-report compared to the interview measure (Kramer et al., 2023). Emphasizing the instruction to anchor responses to the index trauma during administration of the PCL may minimize this issue.

Additionally, differences between trained clinicians and patients in understanding symptom ratings may account for some discrepancy in scoring. Patients may struggle to understand the nuances or accurate definitions of certain symptoms (e.g., flashbacks), resulting in inaccurate overreporting on self-report measures (Kramer et al., 2023). However, in this study, because the CAPS was administered first, participants presumably had the opportunity to seek clarification on these items, which should result in more accurate self-reporting. Despite this, the scoring discrepancies remained. Trained clinicians also have exposure to a wide range of symptom presentations across multiple patients, while participants only have the perspective of their own symptoms. This may also factor into participants rating more extreme levels of distress than the assessors do. This may be an inherent limitation that prevents higher consistency between measurement type.

4.1. Constraints on generality

As with all studies, there are limitations with the present secondary analysis. The sample was comprised of active duty military and veterans, who might respond differently to the assessments than civilians. Because they were being assessed for treatment studies, there was a relatively smaller group without PTSD. There were few women in the sample due to the military and veteran population. When examining diagnostic agreement, we focused on comparing the 20 symptom criteria on the PCL-5 and CAPS-5, but for comparability, we did not include in this analysis other items on the CAPS-5 part of a PTSD diagnosis (e.g. clinician's assessment of functional impairment and duration over a month) that are not reflected in the PCL-5. However, the number who were diagnosed as probable PTSD were nearly identical across the measures using the 20 items on the symptom criteria without the distress/impairment and duration items based on this treatment-seeking sample. The duration and impairment items are probably more important in some other samples. Additionally, in this particular set of studies, participants always received the CAPS-5 first, with its follow-up questions, and yet they still scored higher on the PCL-5, which asked about the same event and symptoms with the same number of questions and scores. It is possible that individuals might be more comfortable writing down the severity of their symptoms than speaking about them aloud to an interviewer. Although the pattern of results observed here is consistent with past research, the lack of counterbalancing may have influenced the results.

Although outcomes of this study match prior research, the smaller sample at posttreatment may reflect bias, and, because the interviewers were trained and monitored by the same research group, clinician training could have been affected. Collecting information from study assessors regarding their perceptions of the delivery and scoring of the interview

measures is a future research direction that may provide additional insight about the scoring discrepancies and highlight considerations for future training. Finally, as with all clinical trials, interviews were closely monitored and supervised for fidelity. As such, the variability in scores identified in this sample may not represent score variability in clinical use. It is recommended that additional studies of the PCL-5 and CAPS-5 with clinical data be conducted.

The strengths of this study were the large sample of treatment-seekers, including those who did not qualify for the respective studies, and a smaller but still substantial sample of participants at a posttreatment assessment. In accordance with best research practice, the CAPS-5 was trained and administered by the same team across studies, and a portion of each study were re-rated for calibration, which was found to be excellent. The PCL-5 was not given in close proximity to the CAPS-5; there was an intentional break between these measures that included a battery of intervening measures that assessed other symptoms or variables of interest (e.g. social support, functioning).

5. Conclusions

The importance of this study is to encourage developers of these and other interview/self-report sets of measures to ensure they are measuring the same diagnostic criteria. It is often not feasible to administer standardized interviews in research projects with large samples or in clinical settings. It becomes unclear what is actually being assessed if the scores between the two assessments vary, resulting in differences between who is included or excluded from research or clinical services. It also makes difficult any comparisons across studies or program evaluation when comparing one type of measure or the other. For instance, in this particular case, the scaling on the PCL-5 appeared to be superior to the CAPS-5 due to the greater use of the full range of scores, yet it is also possible that the language of the questions on the PCL-5 often did not give recipients enough information for them to truly assess the symptom at hand as intended by the *DSM-5*. Ultimately, concordance between the PCL-5 and CAPS-5 is not ideal but could be improved through use of identical rating scale anchors and clarification of self-report questions. Developers and trainers of the CAPS-5 might also consider whether using a wider range of response options should be encouraged. Interviews have been the standard for randomized clinical trials. However, for very large studies, those without funding, and clinical settings, the inexpensive, 5-minute, self-report PCL-5 might be preferable, with confidence, to a 60-minute clinical interview that requires training and fidelity assessments. Overall, given differences in outcomes on the two assessments, researchers and clinicians are encouraged to consider the strengths and weaknesses of both interview and self-report forms as well as the different information each may provide.

Author contributions

Conceptualization: Drs Resick, Straud, Peterson, Young-McCaughan, Mintz. Data curation: Drs Straud, Young-McCaughan, Mintz. Formal analysis: Drs. Straud and Mintz. Funding acquisition: Drs Resick, Peterson, McGeary, Young-McCaughan, and Taylor. Investigation: Drs Resick, LoSavio, Peterson, McGeary, Young-McCaughan, Taylor, and Wachen. Methodology: Drs Resick, Straud, LoSavio, Peterson, McGeary, Young- McCaughan, Taylor, Wachen, Mintz. Project administration: Drs Resick, Peterson, McGeary, Young-McCaughan, Taylor, and Wachen. Resources: Drs Resick, Peterson, McGeary, Young-McCaughan, Taylor, and Wachen. Software: Drs Straud and Mintz. Supervision: Drs Resick, Peterson, McGeary, Young-McCaughan, Taylor, Wachen. Validation: Drs Resick, Straud, and Mintz. Visualization: Drs Straud. Writing – original draft: Drs Resick, Straud, Wachen and LoSavio. Writing – review and editing: Drs Resick, Straud, Wachen, LoSavio, Peterson, McGeary, Young-McCaughan, Taylor, and Mintz.

Clinical trial registration

The parent studies used for these analyses were registered with ClinicalTrials.gov [Identifiers NCT023818, NCT02290847, NCT02419131, and NCT02773693]. This study was not preregistered. The data from this study are maintained at the University of Texas Health Science Center at San Antonio in the STRONG STAR Repository.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

Requests for access to the data as well as for materials and the analysis code also can be emailed to repository@ strongstar.org.

ORCID

Patricia A. Resick http://orcid.org/0000-0002-2203-2404
Casey L. Straud http://orcid.org/0000-0001-8875-1906
Jennifer Schuster Wachen http://orcid.org/0000-0002-3652-5787
Stefanie T. LoSavio http://orcid.org/0000-0003-3653-3194
Alan L. Peterson http://orcid.org/0000-0003-2947-2936
Donald D. McGeary http://orcid.org/0000-0001-9660-2652
Stacey Young-McCaughan http://orcid.org/0000-0002-3327-5233
Daniel J. Taylor http://orcid.org/0000-0003-3433-2516
Jim Mintz http://orcid.org/0000-0002-8299-5851

References

Adkins, J. W., Weathers, F. W., McDevitt-Murphy, M., & Daniels, J. B. (2008). Psychometric properties of seven self-report measures of posttraumatic stress disorder in college students with mixed civilian trauma exposure. *Journal of Anxiety Disorders*, *22*(8), 1393–1402. https:// doi.org/10.1016/j.janxdis.2008.02.002

Allen, M. J., & Yen, W. M. (2002). *Introduction to Measurement Theory*. Waveland Press.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.)

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). https:// doi.org/10.1176/appi.books.9780890425596

Barnes, J. B., Presseau, C., Jordan, A. H., Kline, N. K., Young-McCaughan, S., Keane, T. M., Peterson, A. L., & Litz, B. T. (2019). Common data elements in the assessment of military-related PTSD research applied in the consortium to alleviate PTSD. *Military Medicine*, *184* (5–6), e218–e226. https://doi.org/10.1093/milmed/ usy226

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Klauminzer, G., Charney, D. S., & Keane, T. M. (1990). A clinician rating scale for assessing current and lifetime PTSD: The CAPS-1. *Behavior Therapist*, *13*, 187–188.

Blanchard, E. B., Jones Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy*, *34*(8), 669–673. https://doi.org/10.1016/0005-7967(96)00033-2

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.22059

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

Bollinger, A. R., Cuevas, C. A., Vielhauer, M. J., Morgan, E. E., & Keane, T. M. (2008). The operating characteristics of the PTSD Checklist in detecting PTSD in HIV+ substance abusers. *Journal of Psychological Trauma*, *7*(4), 213–234. https://doi.org/10.1080/19322880802384251

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

Bovin, M. J., & Weathers, F. W. (2022). Assessing PTSD symptoms. In G. Beck & D. Sloan (Eds.), *Oxford Handbook of Traumatic Stress Disorders* (2nd ed, pp. 235–249). Oxford University Press.

Davis, L. W., Schmid, A. A., Daggy, J. K., Yang, Z., O'Connor, C. E., Schalk, N., Do, A. 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*, *12*(8), 904–912. https://doi.org/10.1037/tra0000564

De Jongh, A., Groenland, G. N., Sanches, S., Bongaerts, H., Voorendonk, E. M., & Van Minnen, A. (2020). The impact of brief intensive trauma-focused treatment for PTSD on symptoms of borderline personality disorder. *European Journal of Psychotraumatology*, *11*(1), 1721142. https://doi.org/10.1080/20008198.2020.1721142

Forbes, D., Creamer, M., & Biddle, D. (2001). The validity of the PTSD Checklist as a measure of symptomatic change in combat-related PTSD. *Behaviour Research and Therapy*, *39*(8), 977–986. https://doi.org/10.1016/S0005-7967(00)00084-X

Held, P., Kovacevic, M., Petrey, K., Meade, E. A., Pridgen, S., Montes, M., Werner, B., Miller, M. L., Smith, D. L., Kaysen, D., & Karnik, N. S. (2022). Treating posttraumatic stress disorder at home in a single week using 1-week virtual massed cognitive processing therapy. *Journal of Traumatic Stress*, *35*(4), 1215–1225. https:// doi.org/10.1002/jts.22831

Keen, S. M., Kutter, C. J., Niles, B. L., & Krinsley, K. E. (2008). Psychometric properties of PTSD checklist in sample of male veterans. *Journal of Rehabilitation Research & Development*, *45*(3), 465–474. https://doi. org/10.1682/JRRD.2007.09.0138

Kramer, L. B., Whiteman, S. E., Petri, J. M., Spitzer, E. G., & Weathers, F. W. (2023). Self-rated versus clinician-rated assessment of posttraumatic stress disorder: An evaluation of discrepancies between the PTSD Checklist for DSM-5 and the Clinician-Administered PTSD Scale for DSM-5. *Assessment*, *30*(5), 1590–1605. https://doi.org/ 10.1177/10731911221113571

Krüger-Gottschalk, A., Ehring, T., Knaevelsrud, C., Dyer, A., Schäfer, I., Schellong, J., Rau, H., & Köhler, K. (2022).

Confirmatory factor analysis of the clinician-administered PTSD scale (CAPS-5) based on DSM-5 vs. ICD-11 criteria. *European Journal of Psychotraumatology*, 13(1), 2010995. https://doi.org/10.1080/20008198.2021.2010995

Lee, D. J., Weathers, F. W., Thompson-Hollands, J., Sloan, D. M., & Marx, B. P. (2022). Concordance in PTSD symptom change between DSM-5 versions of the Clinician-Administered PTSD Scale and the PTSD Checklist (PCL-5). *Psychological Assessment*, 34(6), 604–609. https://doi.org/10.1037/pas0001130

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Livingston, N. A., Brief, D. J., Miller, M. W., & Keane, T. M. (2021). Assessment of PTSD and its comorbidities in adults chapter 16. In M. J. Friedman, P. P. Schnurr, & T. M. Keane (Eds.), *Handbook of PTSD: Science and practice* (3rd ed., pp. 283–298). The Guilford Press.

McGeary, D. D., Resick, P. A., Penzien, D. B., McGeary, C. A., Houle, T. T., Eapen, B. C., Jaramillo, C. A., Nabity, P. S., Reed, D. E. I. I., Moring, J. C., Bira, L. M., Hansen, H. R., Young-McCaughan, S., Cobos, B. A., Mintz, J., Keane, T. M., & Peterson, A. L. (2022). Cognitive behavioral therapy for veterans with comorbid posttraumatic headache and posttraumatic stress disorder symptoms: A randomized clinical trial. *JAMA Neurology*, 79(8), 746–757. https://doi.org/10.1001/jamaneurol.2022.1567

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22, 276–282. https://doi.org/10.11613/BM.2012.031

Monson, C. M., Gradus, J. L., Young-Xu, Y., Schnurr, P. P., Price, J. L., & Schumm, J. A. (2008). Change in posttraumatic stress disorder symptoms: Do clinicians and patients agree? *Psychological Assessment*, 20(2), 131–138. https://doi.org/10.1037/1040-3590.20.2.131

Peterson, A. L., Mintz, J., Moring, J. C., Straud, C. L., Young-McCaughan, S., McGeary, C. A., McGeary, D. D., Litz, B. T., Velligan, D. I., Macdonald, A., Mata-Galan, E., Holliday, S. L., Dillon, K. H., Roache, J. D., Bira, L. M., Nabity, P. S., Medellin, E. M., Hale, W. J., & Resick, P. (2022). In-office, in-home, and telehealth cognitive processing therapy for posttraumatic stress disorder in veterans: A randomized clinical trial. *BMC Psychiatry*, 22(1), 41. https://doi.org/10.1186/s12888-022-03699-4.

Peterson, A. L., Young-McCaughan, S., Roache, J. D., Mintz, J., Litz, B. T., Williamson, D. E., Resick, P. A., Foa, E. B., McGeary, D. D., Dondanville, K. A., Taylor, D. J., Wachen, J. S., Fox, P. T., Bryan, C. J., McLean, C. P., Pruiksma, K. E., Yarvis, J. S., Niles, B. L., Abdallah, C. G., … for the STRONG STAR Consortium and the Consortium to Alleviate PTSD. (2021). STRONG STAR and the Consortium to Alleviate PTSD: Shaping the future of combat PTSD and related conditions in military and veteran populations. *Contemporary Clinical Trials*, 110, Article 106583. https://doi.org/10.1016/j.cct.2021.106583

Resick, P. A., Wachen, J. S., Dondanville, K. A., LoSavio, S. T., Young-McCaughan, S., Yarvis, J. S., Pruiksma, K. E.,

Blankenship, A., Jacoby, V., Peterson, A. L., Mintz, J., & for the STRONG STAR Consortium. (2021). Variable-length cognitive processing therapy for posttraumatic stress disorder in active duty military: Outcomes and predictors. *Behaviour Research and Therapy*, 141, Article 103846. https://doi.org/10.1016/j.brat.2021.103846

Stein, N. R., Mills, M. A., Arditte, K., Mendoza, C., Borah, A. M., Resick, P. A., Litz, B. T., & STRONG STAR Consortium. (2012). A scheme for categorizing traumatic military events. *Behavior Modification*, 36(6), 787–807. https://doi.org/10.1177/0145445512446945

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Taylor, D. J., Pruiksma, K. E., Mintz, J., Slavish, D. C., Wardle-Pinkston, S., Dietch, J. R., Dondanville, K. A., Young-McCaughan, S., Nicholson, K. L., Litz, B. T., Keane, T. M., Peterson, A. L., Resick, P. A., & on behalf of the Consortium to Alleviate PTSD. (2022). *Treatment of Comorbid Sleep Disorders and Posttraumatic Stress Disorder in U.S. Active Duty Military Personnel: A Pilot Study* [Manuscript Submitted for Publication]. Department of Psychology, University of Arizona.

Vogt, D. S., Smith, B. N., King, L. A., King, D. W., Knight, J. A., & Vasterling, J. J. (2013). Deployment risk and resilience inventory-2 (DRRI-2): An updated tool for assessing psychosocial risk and resilience factors among service members and veterans. *Journal of Traumatic Stress*, 26(6), 710–717. https://doi.org/10.1002/jts.21868

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013a). *The clinician-administered PTSD scale for DSM-5 (CAPS-5)* [Assessment]. U.S. Department of Veterans Affairs, National Center for PTSD. https://www.ptsd.va.gov/professional/assessment/adult-int/caps.asp.

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013b). *The life events checklist for DSM-5 (LEC-5)* [Assessment]. U.S. Department of Veterans Affairs, National Center for PTSD. https://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp.

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., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD checklist (PCL): Reliability, validity, and diagnostic utility* [Paper presentation]. 9th annual Meeting of the International Society for Traumatic Stress Studies, San Antonio, TX, United States.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD checklist for DSM-5 (PCL-5)* [Assessment]. U.S. Department of Veterans Affairs, National Center for PTSD. https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp.