Psychological Assessment
2016, Vol. 28, No. 11, 1392–1403

http://dx.doi.org/10.1037/pas0000260

Psychometric Analysis of the PTSD Checklist-5 (PCL-5) Among Treatment-Seeking Military Service Members

Jennifer H. Wortmann and Alexander H. Jordan
VA Boston Healthcare System, Jamaica Plain, Massachusetts, and Boston University School of Medicine

Frank W. Weathers
Auburn University

Patricia A. Resick
Duke University Medical Center

Katherine A. Dondanville and Brittany Hall-Clark
The University of Texas Health Science Center at San Antonio

Edna B. Foa
University of Pennsylvania

Stacey Young-McCaughan
The University of Texas Health Science Center at San Antonio

Jeffrey S. Yarvis
Carl R. Darnall Army Medical Center, Fort Hood, Texas

Elizabeth A. Hembree
University of Pennsylvania

Jim Mintz
The University of Texas Health Science Center at San Antonio

Alan L. Peterson
The University of Texas Health Science Center at San Antonio and South Texas Veterans Health Care System, San Antonio, Texas

Brett T. Litz
VA Boston Healthcare System, Jamaica Plain, Massachusetts, and Boston University School of Medicine

On behalf of the STRONG STAR Consortium

The Posttraumatic Stress Disorder Checklist (PCL-5; Weathers et al., 2013) was recently revised to reflect the changed diagnostic criteria for posttraumatic stress disorder (PTSD) in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–5*; American Psychiatric Association, 2013). We investigated the psychometric properties of PCL-5 scores in a large cohort (N = 912) of

This article was published Online First January 11, 2016.

Jennifer H. Wortmann and Alexander H. Jordan, Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, Jamaica Plain, Massachusetts, and Department of Psychiatry, Boston University School of Medicine; Frank W. Weathers, Department of Psychology, Auburn University; Patricia A. Resick, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center; Katherine A. Dondanville and Brittany Hall-Clark, Department of Psychiatry, The University of Texas Health Science Center at San Antonio; Edna B. Foa, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania; Stacey Young-McCaughan, Department of Psychiatry, The University of Texas Health Science Center at San Antonio; Jeffrey S. Yarvis, Department of Behavioral Health, Carl R. Darnall Army Medical Center, Fort Hood, Texas; Elizabeth A. Hembree, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania; Jim Mintz, Department of Psychiatry and Department of Epidemiology & Biostatistics, The University of Texas Health Science Center at San Antonio; Alan L. Peterson, Department of Psychiatry, The University of Texas Health Science Center at San Antonio, and Office of Research and Development, South Texas Veterans Health Care System, San Antonio, Texas; Brett T. Litz, Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, and Department of Psychiatry, Boston University School of Medicine; on behalf of the STRONG STAR Consortium.

This material is the result of work supported with resources and the use of facilities at the VA Boston Healthcare System and South Texas Veterans Health Care System. The views expressed in this article are solely those of the authors and do not reflect an endorsement by or the official policy of the U.S. Army, the Department of Defense, the Department of Veterans Affairs, or the U.S. Government. Dr. Foa receives royalties from the sale of *Prolonged Exposure Therapy for PTSD: Emotional Processing of Traumatic Experiences Therapist Guide* and *Reclaiming Your Life From a Traumatic Experience Workbook* by Oxford University Press (New York, NY). Dr. Foa also receives payment for training she conducts on prolonged exposure therapy. Funding for this work was made possible by the U.S. Department of Defense through the U.S. Army Medical Research and Materiel Command, Congressionally Directed Medical Research Programs, Psychological Health and Traumatic Brain Injury Research Program awards W81XWH-08-02-109 (Alan Peterson), W81XWH-08-02-0111 (Edna Foa), W81XWH-08-02-0114 (Brett Litz), and W81XWH-08-02-0116 (Patricia Resick). We thank Robert A. Lew, PhD, for his consultation on the clinical change analyses.

Correspondence concerning this article should be addressed to Jennifer H. Wortmann, VA Boston Healthcare System, 150 South Huntington Avenue (151-MAV), Boston, MA 02130. E-mail: jennifer.wortmann@va.gov

military service members seeking PTSD treatment while stationed in garrison. We examined the internal consistency, convergent and discriminant validity, and *DSM–5* factor structure of PCL-5 scores, their sensitivity to clinical change relative to PTSD Symptom Scale–Interview (PSS-I; Foa, Riggs, Dancu, & Rothbaum, 1993) scores, and their diagnostic utility for predicting a PTSD diagnosis based on various measures and scoring rules. PCL-5 scores exhibited high internal consistency. There was strong agreement between the order of hypothesized and observed correlations among PCL-5 and criterion measure scores. The best-fitting structural model was a 7-factor hybrid model (Armour et al., 2015), which demonstrated closer fit than all other models evaluated, including the *DSM–5* model. The PCL-5's sensitivity to clinical change, pre- to posttreatment, was comparable with that of the PSS-I. Optimally efficient cut scores for predicting PTSD diagnosis were consistent with prior research with service members (Hoge, Riviere, Wilk, Herrell, & Weathers, 2014). The results indicate that the PCL-5 is a psychometrically sound measure of *DSM–5* PTSD symptoms that is useful for identifying provisional PTSD diagnostic status, quantifying PTSD symptom severity, and detecting clinical change over time in PTSD symptoms among service members seeking treatment.

Keywords: PTSD, PCL-5, military, psychometrics

Supplemental materials: http://dx.doi.org/10.1037/pas0000260.supp

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–5*; American Psychiatric Association, 2013) brought major changes to the criteria for posttraumatic stress disorder (PTSD). PTSD was moved from the anxiety disorders category to a new class of "trauma- and stressor-related disorders," and Criterion A2 (peritraumatic fear, helplessness, or horror) was eliminated. The three symptom clusters in the fourth edition of the *DSM* (*DSM–IV*; American Psychiatric Association, 1994) were expanded into four by splitting the avoidance and numbing cluster into an *avoidance* cluster and a *negative alterations in cognitions and mood* cluster. Three new symptoms were introduced (persistent and distorted blame of self or others, persistent negative emotional state, and reckless or self-destructive behavior), and several others were revised. Also, a dissociative PTSD subtype was added, as was a requirement that symptoms not be attributable to medications, substance use, or other illness.

Clinical interviews and self-report instruments for assessing PTSD have been revised to align with the *DSM–5* changes, and the psychometric characteristics of these new tests need to be evaluated (Weathers, Marx, Friedman, & Schnurr, 2014). One of the most widely used self-report questionnaires, especially in military and veteran contexts, is the *DSM–IV*-based PTSD Checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993). In this report, we describe the psychometric properties of the PCL for *DSM–5* (PCL-5; Weathers et al., 2013, available at http://www.ptsd.va.gov/) in a cohort of service members seeking treatment for deployment-related PTSD. Our use of a treatment-seeking military sample makes the results especially germane to researchers and clinicians studying and treating service members and new veterans.

Psychometric evaluations of PCL-5 scores to date have found them to have high internal consistency in samples of war veterans ($\alpha = .95$; Pietrzak et al., 2015) as well as civilians ($\alpha = .95$; Armour et al., 2015). Evidence for construct validity has likewise been encouraging but limited in scope. In a large cohort of U.S. Army Soldiers, total PCL-5 scores were associated with reports of depression ($r = .73$) and generalized anxiety ($r = .79$) as well as greater functional impairment (rs ranging from .31 to .59; Hoge, Riviere, Wilk, Herrell, & Weathers, 2014). In a civilian sample of undergraduates with exposure to potentially traumatic events, scores on the four *DSM–5* subcomponents of the PCL-5 were associated with depressive symptoms (rs ranging from .48 to .78; Biehn et al., 2013). Other studies evaluating the convergent and discriminant validity of PCL-5 scores have used factor structures deviating from the *DSM–5* (Pietrzak et al., 2015; Tsai et al., 2015), rendering unclear what can be generalized from these efforts.

Most prior investigations of the PCL-5 have focused on testing the validity of the *DSM–5* factor structure, with highly variable results. A study of war veterans enrolled in an online intervention reported adequate fit and temporal stability of the *DSM–5* structure (Keane et al., 2014). A study with undergraduate students also found support for the *DSM–5* model compared with a four-factor model that combines emotional numbing and hyperarousal into a Dysphoria factor (Biehn et al., 2013). On the other hand, in a sample of Chinese earthquake survivors, a six-factor model—dividing Criterion D into negative affect and anhedonia components, and Criterion E into anxious arousal and dysphoric arousal components—best fit the data (Liu et al., 2014). In addition, analyses of a large, nationally representative sample of veterans found support for a seven-factor model that includes negative affect, anhedonia, anxious arousal, dysphoric arousal, and externalizing behaviors symptom clusters (Armour et al., 2015; Pietrzak et al., 2015).

In the only published study examining the clinical impact and utility of the PCL-5 in military personnel, Hoge and colleagues (2014) compared the PCL-5 with the *DSM–IV* specific stressor version of the PCL (PCL-S; Weathers et al., 1993) in a large sample of U.S. Army Soldiers. Although the PCL-5 and PCL-S showed a high degree of agreement in identifying PTSD cases ($\kappa = .67$), Hoge et al. found substantial discordance between the measures; 30% of Soldiers who met criteria for PTSD on the PCL-S did not meet criteria on the PCL-5, and 27% who met criteria on the PCL-5 did not meet criteria on the PCL-S. The investigators also identified optimally efficient cut scores on the PCL-5 that corresponded to empirically derived cut scores on the PCL-S indicating a positive screen for PTSD.

Prior evaluations of the PCL-5 have thus focused on factor structure or comparison of diagnostic criteria and have not comprehensively analyzed the psychometric characteristics of the measure. We conducted a psychometric evaluation of PCL-5 scores in

a sample of service members seeking treatment for deployment-related PTSD in the context of randomized controlled psychotherapy trials within the South Texas Research Organizational Network Guiding Studies on Trauma and Resilience (STRONG STAR) consortium. We evaluated the internal consistency, convergent and discriminant validity, and structural validity of PCL-5 scores. Because most participants in our cohort were treated for PTSD, we also examined the extent to which PCL-5 scores indexed change. In addition, to assist clinicians and researchers who plan to use the PCL-5 as a repeated measure of change, we derived PCL-5 cut scores that would indicate clinically meaningful symptom improvement or exacerbation among service members treated for PTSD. Finally, we calculated optimally efficient cut scores on the PCL-5 that corresponded to diagnostic cut scores on a *DSM–IV*-based clinical interview and the PCL-S.

Method

Participants and Procedure

Participants were military service members and recently retired veterans recruited, assessed, and treated at the Carl R. Darnall Army Medical Center at Fort Hood, Texas, under the auspices of the STRONG STAR Consortium. We combined data from three clinical trials in the consortium. Recruitment was based on referrals by providers and self-referrals by service members. Prescreening criteria included active duty service member status, deployed in support of Iraq and/or Afghanistan wars, aged 18 to 65 years, willing to participate in the format of treatments used in the research, available for the duration of the study, and no recent medication changes. To ensure the presence of noncases in our sample, we included participants ($n = 147$) who completed baseline measures after consenting to participate but were not randomized into a trial (and thus did not complete follow-up measures) because they did not meet diagnostic criteria for PTSD on the PTSD Symptom Scale–Interview version (PSS-I; Foa, Riggs, Dancu, & Rothbaum, 1993). The results of this study will therefore be chiefly generalizable to treatment-seeking military personnel. Participants were not compensated for their involvement in the study.

The study was approved by the institutional review boards of the STRONG STAR collaborating institutions. At baseline, participants completed a battery of 39 measures that included the PCL-S (19th in order) and the PCL-5 (39th), and they were administered the PSS-I by specially trained independent (blinded) evaluators (16th). Participants selected their most currently distressing Criterion A event before administration of the PSS-I, and this event also served as the index event for the PCL-S and the PCL-5. To identify possible index events, participants first completed the Life Events Checklist (Gray, Litz, Hsu, & Lombardo, 2004) and modified versions (see the Measures section) of the Combat Experiences and Aftermath of Battle subscales of the Deployment Risk and Resilience Inventory (DRRI; King, King, Vogt, Knight, & Samper, 2006). Independent evaluators guided each participant in selecting the most distressing event of those identified on these forms. For service members randomized into a clinical trial, the same instruments were also administered in the same order 2 weeks after the end of treatment. A sample of 912 participants completed the PCL-5 as well as the PCL-S and PSS-I at baseline. Additional measures that were used to assess discriminant validity of PCL-5 scores were available for subsamples of these participants.

Measures

Demographics. A demographics form measured race, gender, age, education, military service information (e.g., military grade), and other participant characteristics (see Table 1).

PCL-5. The PCL-5 is a 20-item self-report measure that evaluates the degree to which an individual has been bothered in the past month by *DSM–5* PTSD symptoms tied to his or her most currently distressing event (Weathers et al., 2013). Items are rated from 0 (*not at all*) to 4 (*extremely*) and are summed for a total severity score. Subscale severity scores are calculated by summing items in each of the four *DSM–5* PTSD symptom clusters: intrusions (Items 1–5), avoidance (Items 6–7), negative alterations in cognitions and mood (NACM; Items 8–14), and alterations in

Table 1
Participant Characteristics

Variable	*M* (*SD*)
Age ($n = 909$)	32.67 (7.43)
Months in military ($n = 912$)	128.81 (77.25)
Months since return from last deployment ($n = 768$)	23.29 (21.51)
	% (*n*)
Gender	100 (912)
Male	91.6 (835)
Female	8.4 (77)
Race	99.1 (904)
American Indian/Alaska Native	2.0 (18)
Asian	1.2 (11)
Native Hawaiian or Other Pacific Islander	1.4 (13)
Black or African American	24.5 (223)
White	57.0 (520)
Other	13.0 (119)
Ethnicity	99.8 (910)
Hispanic or Latino	19.2 (175)
Not Hispanic or Latino	80.6 (735)
Education	99.9 (911)
Some high school	.4 (4)
GED	7.9 (72)
High school diploma	23.1 (211)
Some college	50.5 (461)
Associate degree	9.8 (89)
4-year college degree or higher	8.1 (74)
Marital status: % married or living with a partner	73.8 (673)
Military status	100 (912)
Active Duty	98.6 (899)
Reserve	.7 (6)
National Guard	.5 (5)
Retired	.2 (2)
Enlisted (versus Officer)	96.9 (884)
Deployed to Iraq or Afghanistan	99.9 (911)
One time	29.3 (267)
Two times	32.8 (299)
Three times	23.1 (211)
Four or more times	14.6 (133)
Military role	100 (912)
Combat arms	42.2 (385)
Combat support	22.0 (201)
Combat service support	35.7 (326)

Note. Total sample size is 912. Because of incomplete data for some participant characteristics, specific *n* values are noted with each variable.

arousal and reactivity (AR; Items 15–20). PTSD caseness was defined as endorsing a severity of at least a 2 (*moderate*) for a sufficient number of symptoms in each cluster to meet *DSM–5* criteria. At the time this study began, only a draft version of the PCL-5 was available. Minor wording differences between the version of the PCL-5 used in this study and the currently published version are described in the online supplemental materials. Internal consistency reliabilities in this study for the PCL-5 and all measures enumerated in the following section are shown in Table 2.

Measures Used to Assess Convergent and Discriminant Validity

PCL-S. The PCL-S is a 17-item self-report measure that evaluates the degree to which an individual has been bothered in the past month by *DSM–IV* PTSD symptoms tied to a specific life event (Weathers et al., 1993). Items are rated from 1 (*not at all*) to 5 (*extremely*) and are summed for a total severity score. Subscale severity scores are calculated by summing items in each of the three *DSM–IV* PTSD symptom clusters: intrusions (Items 1–5), avoidance/emotional numbing (Items 6–12), and arousal (Items 12–17). Additionally, to permit comparison with the PCL-5 items capturing the revised *DSM–5* Criterion C, a strategic avoidance subscale was calculated by summing Items 6 and 7. PCL-S scores have shown high internal consistency and strong convergent validity in military samples (Wilkins, Lang, & Norman, 2011). Caseness on the PCL-S was defined as endorsing a severity of at least a 3 (*moderate*) for a sufficient number of symptoms in each cluster to meet *DSM–IV* criteria. Of the three versions of the PCL developed for *DSM–IV*, the PCL-S and PCL-5 most closely resemble each other because they contain an instructional prompt to respond to items in the context of a specific life event. The PCL-S was used rather than the PCL-M (military version) because the PCL-S requires the endorsement of symptoms tied to a specific life event. The PCL-5 and PCL-S measures share similar instructions and contain overlapping items, with the exceptions described in the introduction. Hoge and colleagues (2014) provide a detailed item-by-item comparison of the PCL-5 and PCL-S.

PSS-I. The PSS-I is a 17-item clinical interview that evaluates *DSM–IV* PTSD symptoms on a single dimension combining frequency and severity in the last 2 weeks (Foa et al., 1993). PSS-I scores have been shown to be highly correlated with scores on the Clinician Administered PTSD Scale (Blake et al., 1995), which takes longer to administer (Foa & Tolin, 2000). Items are scored on a scale ranging from 0 (*not at all*) to 3 (*very much*) and are

Table 2
Descriptive Statistics for Study Measures

Variable (possible range)	Baseline			Follow-up (2 weeks posttreatment)		
	M (*SD*)	*n*	Cronbach's alpha	*M* (*SD*)	*n*	Cronbach's alpha
PCL-5 total (0–80)	42.41 (15.06)	912	.91	34.08 (19.01)	439	.95
Intrusions (0–20)	10.56 (4.45)		.80	7.90 (5.55)		.92
Avoidance (0–8)	4.88 (2.35)		.83	3.53 (2.65)		.92
NACM (0–28)	12.91 (6.59)		.82	10.68 (7.49)		.89
AR (0–24)	14.07 (4.68)		.75	11.97 (5.61)		.84
PCL-S total (17–85)	53.11 (12.27)	912	.88	44.68 (16.21)	439	.94
Intrusions (5–25)	15.09 (4.29)		.80	12.20 (5.25)		.91
Avoidance/Emotional numbing (7–35)	20.08 (6.13)		.81	16.92 (7.32)		.89
Strategic avoidance (2–10)	6.77 (2.24)		.74	5.20 (2.67)		.89
Arousal (5–25)	17.94 (4.20)		.75	15.55 (5.28)		.84
PSS-I total (0–51)	22.91 (8.24)	912	.79	19.51 (9.52)	439	.85
Intrusions (0–15)	4.62 (2.98)		.60	3.55 (3.08)		.68
Avoidance/Emotional numbing (0–21)	8.40 (4.15)		.69	6.79 (4.68)		.76
Strategic avoidance (0–6)	3.50 (1.84)		.40	2.61 (2.20)		.68
Arousal (0–15)	9.89 (3.16)		.54	9.16 (3.54)		.60
DRRI Combat experiences (23–115)	49.42 (15.39)	907	.91	N/A		N/A
DRRI Aftermath of battle (15–75)	33.66 (11.56)	910	.92	N/A		N/A
BAI (0–63)	22.91 (12.35)	909	.92	18.33 (12.75)	474	.94
BDI (0–63)	26.55 (11.16)	908	.92	20.83 (13.25)	475	.95
TRGI: Hindsight-bias/Responsibility (0–28)	6.76 (7.31)	910	.88	5.67 (6.89)	464	.91
TRGI: Wrongdoing (0–20)	5.13 (4.64)	910	.73	5.09 (4.85)	464	.79
TRGI: Lack of justification (0–16)	7.65 (5.06)	910	.82	6.96 (4.99)	464	.85
STAXI-2 S-Anger (0–60)	23.25 (11.02)	911	.96	21.70 (10.74)	469	.96
ISI (0–28)	17.84 (5.55)	513	.80	17.23 (6.43)	152	.87
AUDIT (0–40)	4.85 (5.41)	911	.76	3.65 (4.59)	363	.70
PHQ-15 (0–30)	12.55 (4.75)	768	.81	11.28 (5.27)	378	.80
RSES (0–88)	53.46 (16.78)	910	.93	51.05 (17.41)	225	.94

Note. Both groups include noncases. PCL-5 = PTSD Checklist for *DSM-5*; *DSM-5* = fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); NACM = Negative Alterations in Cognitions and Mood; AR = Alterations in Arousal and Reactivity; PCL-S = PTSD Checklist specific stressor version for *DSM-IV*; *DSM-IV* = fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994); PSS-I = PTSD Symptom Scale–Interview version; DRRI = Deployment Risk and Resilience Inventory; BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; TRGI = Trauma-Related Guilt Inventory; STAXI-2 S-Anger = State-Trait Anger Expression Inventory-2, State Anger subscale; ISI = Insomnia Severity Index; AUDIT = Alcohol Use Disorders Identification Test; PHQ-15 = Patient Health Questionnaire-15; RSES = Response to Stressful Experiences Scale; N/A = Not applicable.

summed for a total severity score. Subscale severity scores are calculated by summing items in each of the *DSM–IV* PTSD symptom clusters, as described in this section for the PCL-S. Additionally, to permit comparison with the PCL-5 items capturing the revised *DSM–5* Criterion C, a strategic avoidance subscale was calculated by summing Items 6 and 7. PSS-I scores have shown high internal consistency and strong convergent validity in civilian samples (Foa et al., 1993; Foa & Tolin, 2000). PTSD caseness on the PSS-I was calculated in two ways. First, as recommended by the scale developers, symptoms were rated as present if they were scored at least a 1 (*a little*), and caseness was defined as the presence of a sufficient number of symptoms in each cluster to meet *DSM–IV* criteria (Foa et al., 1993). Second, we generated a more stringent case definition, matching the PCL-based decision rule, requiring a severity of at least a 2 (*somewhat*) for the requisite number of symptoms to meet *DSM–IV* criteria.

Warzone exposure. The Combat Experiences and Aftermath of Battle subscales of the DRRI (King et al., 2006) were modified so that the items asked only about whether the individual faced exposure to various stressors himself or herself, rather than asking whether the individual or members of his or her unit had been exposed. Additionally, eight items were added to the Combat subscale, assessing events likely to have been encountered in the Iraq and Afghanistan wars (see Vasterling et al., 2010). Responses to all items were given on a scale ranging from 1 (*never*) to 5 (*daily or almost daily*), and scores were summed to create total Combat Experiences and Aftermath of Battle severity scores (see Vogt, Proctor, King, King, & Vasterling, 2008 and Vogt, Smith, King, & King, 2012). Scores on these subscales have shown high internal consistency and strong concurrent and discriminant validity in veterans of the war in Iraq (Vogt et al., 2008).

Anxiety symptoms. The Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988) is a 21-item self-report measure that asks respondents to rate the extent to which they have been bothered by anxiety symptoms within the past week on a scale ranging from 0 (*not at all*) to 3 (*severely*). Scores are summed to obtain a total severity score. BAI scores have shown high internal consistency and good convergent and discriminant validity in civilian samples (Beck et al., 1988).

Depression symptoms. The Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996) is a 21-item self-report measure that assesses symptoms of depression. For each item, respondents choose one of four statements representing various levels of symptom severity over the past 2 weeks; the statements are scaled from 0 (*no disturbance*) to 3 (*maximal disturbance*). Scores are summed to obtain a total severity score. BDI-II scores have shown high internal consistency and good convergent and discriminant validity in civilian samples (Beck et al., 1996).

Guilt. The Trauma-Related Guilt Inventory (TRGI) is a self-report measure of guilt-related cognitions and distress (Kubany et al., 1996). Whereas the original scale includes 32 items tied to a specific traumatic experience, we used an abbreviated 16-item version tied to combat and operational experiences generally. Items are rated from 0 (*not at all true*) to 4 (*extremely true*) and summed into three subscales: Hindsight-Bias/Responsibility, Wrongdoing, and Lack of Justification. Scores on the TRGI subscales have shown moderate to high internal consistency, strong convergent validity with trait measures of guilt, and good discriminant validity with age and education in a sample of Vietnam veterans (Kubany et al., 1996).

Anger. The State-Trait Anger Expression Inventory-2 (STAXI-2; Spielberger, 1999) State Anger subscale is a 15-item self-report measure of the extent to which an individual currently feels or wishes to express anger. Items are rated on a scale ranging from 1 (*not at all*) to 4 (*very much so*) and are summed to create an overall intensity score. Scores on the STAXI-2 have shown high internal consistency and strong convergent validity in civilian samples (Spielberger, 1999), and scores on the original STAXI have shown high internal consistency in a Vietnam veteran sample (Taft, Street, Marshall, Dowdall, & Riggs, 2007).

Sleep disturbance. The Insomnia Severity Index (ISI; Morin, 1993) is a seven-item self-report measure of perceived insomnia severity. Items are scaled from 0 (*no disturbance*) to 4 (*maximal disturbance*) and summed to create a total severity score. ISI scores have shown high internal consistency among veterans of Operation Enduring Freedom/Operation Iraqi Freedom (OEF/OIF; Epstein, Goren, & Bushnell, 2009).

Alcohol use. The Alcohol Use Disorders Identification Test (AUDIT; Babor, Higgins-Biddle, Saunders, & Monteiro, 2001) is a 10-item screening measure that assesses alcohol consumption, drinking behavior, and alcohol-related problems. Items are rated on a 4-point scale and summed to create a total severity score. The AUDIT was administered as an interview, with an assessment timeframe of the past 3 months. AUDIT scores (self-report versions) have shown high internal consistency and good diagnostic utility in samples of older veterans (Reinert & Allen, 2007) and OEF/OIF veterans (e.g., Crawford et al., 2013).

Somatic symptoms. The Patient Health Questionnaire–15 (PHQ-15; Kroenke, Spitzer, & Williams, 2002) is a 15-item self-report measure that assesses the degree to which individuals have been bothered by somatic complaints over the past 4 weeks. Items are rated on a scale from 0 (*not bothered at all*) to 2 (*bothered a lot*) and summed for a total severity score. PHQ-15 scores have shown high internal consistency in samples of civilians (Kroenke et al., 2002) and OEF/OIF veterans (Afari et al., 2015).

Resilience. The Response to Stressful Experiences Scale (RSES), developed by National Center for PTSD experts, is a 22-item self-report measure of cognitive, emotional, and behavioral aspects of trait resilience (Johnson et al., 2011). Respondents rate, on a 0 (*not at all like me*) to 4 (*exactly like me*) scale, how well each statement describes them. Items are summed to create a total score. RSES scores have shown high internal consistency and good convergent validity with another measure of resilience in a sample of service members and veterans (Johnson et al., 2011).

Data Analysis Plan and Predictions

Convergent and discriminant validity. We calculated zero-order correlations to evaluate convergent relationships between the PCL-5 and other PTSD measures and discriminant relationships between the PCL-5 and all other measures. Following Westen and Rosenthal (2003), we examined the extent to which the observed pattern of correlations matched the pattern of correlations predicted by prior research. We based our predictions on the range of correlations found in prior literature involving the *DSM–IV* version of the PCL (see the Measures section). We predicted the largest correlation between PCL-5 and PCL-S scores, as both are self-

report measures of PTSD, and a slightly smaller correlation between PCL-5 and PSS-I scores. We predicted that PCL-5 scores would show large correlations with scores on the BAI, BDI-II, ISI, and PHQ-15; moderate correlations with the DRRI Combat Experiences and Aftermath of Battle, TRGI Hindsight-Bias/Responsibility and Wrongdoing, STAXI-2 Anger, and RSES scores (with an inverse relationship for the RSES); small correlations with the AUDIT scores; and no relationship with the TRGI Lack of Justification subscale scores. We calculated two effect size estimates of the fit between predictions and observations: first, $r_{\text{alerting-CV}}$, which reflects the correspondence between the order of the predicted versus observed correlations; and second, $r_{\text{contrast-CV}}$, which accounts for sample size, median intercorrelations among criterion measures, and the magnitudes of correlations between the PCL-5 and criterion measures.

Structural validity. Using Mplus version 7.11 (Muthén & Muthén, 2013), we employed confirmatory factor analysis (CFA) to test the fit of the four-factor *DSM–5* model, using pretreatment data. We also examined the structural validity of the PCL-5 by comparing the *DSM–5* structure with four alternative factor structures for which there is precedent in the PTSD literature (see the online supplemental materials). The five-factor dysphoric arousal model (Elhai et al., 2011) labels NACM "numbing" and divides AR into dysphoric arousal and anxious arousal. The six-factor externalizing behaviors model further divides dysphoric arousal into externalizing behaviors and dysphoric arousal; the six-factor anhedonia model maintains anxious and dysphoric arousal but separates avoidance into negative affect and anhedonia; and the seven-factor hybrid model blends the two six-factor models and comprises intrusions, avoidance, negative affect, anhedonia, externalizing behaviors, anxious arousal, and dysphoric arousal (Armour et al., 2015).

For the CFAs, items were specified to load on only one factor, latent factors were allowed to correlate, and error covariances were set to zero. The first item in each factor was set equal to 1 to identify the model, with the exception of the third factor. We selected Item 11 as the marker variable per Keane and colleagues (2014), because Item 8 (psychogenic amnesia) tends to load poorly on this factor (Miller et al., 2013). Model fit was determined using the following goodness-of-fit indices: chi-square, root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker Lewis index (TLI), and standardized root-mean-square residual (SRMR). CFI and TLI scores above .95, RMSEA values less than .06, and SRMR values less than .08 were interpreted to indicate close fit (Hu & Bentler, 1999). Relative model fit was assessed with chi-square difference tests for nested models and the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for non-nested models. Differences of greater than 10 points between AIC and BIC values were interpreted to indicate that the model with the lower value more closely fit the data (Burnham & Anderson, 2004).

Sensitivity to clinical change. We conducted two sets of analyses to assess the PCL-5 scores' sensitivity to clinical change, relative to the PSS-I, using completers from a single STRONG STAR clinical trial ($N = 91$) to minimize error related to multiple treatment conditions and time durations of the different trials' therapies. To correct for the different response scales on the two measures (0 to 3 for the PSS-I, and 0 to 4 for the PCL-5), we rescaled the PSS-I responses by multiplying values by 4/3. Thus, total scores for the 17 items shared by both instruments could vary from 0 to 68.

First, we assessed pairwise agreement between the PCL-5 and PSS-I in identifying various magnitudes of change using arbitrary cut points. We chose change cut points of 5, 10, 15, and 20 points for both scales and categorized participants as improved or not improved (i.e., the score decreased by the cut point, or did not). Then, we calculated the agreement between the scales, within categorizations, and calculated the percentage of cases in agreement. Cases in agreement were defined as those for which changes on the PCL-5 and PSS-I were either both above or both below the cut point. In addition, we calculated the Pearson correlation and concordance correlation coefficients of pre- and posttreatment PCL-5 and PSS-I change scores, the latter reflecting the extent to which the change scores were of equal magnitude.

Second, we assessed pairwise agreement between the PCL-5 and PSS-I in classifying individuals according to strict criteria for defining statistically reliable and clinically meaningful change. Following Jacobson and Truax's (1991) method, we established reasonable cut scores between the dysfunctional and functional populations using their suggested Cut Score A, defined as the point that is two standard deviations beyond the range of the pretherapy mean. Next, we calculated a reliable change (RC) index for each participant to ensure that changes were not attributable to chance or measurement error. The RC is computed as $(x_2 - x_1)/S_{\text{diff}}$, where x_1 represents the participant's pretreatment PCL-5 or PSS-I total score, x_2 represents the participant's posttreatment total score, and S_{diff} is the standard error of difference between the two test scores. S_{diff} was calculated based on test-retest reliabilities (r_{xx}) available for the measures. For the PCL-5, 30-day $r_{xx} = .84$ (Bovin et al., in press), and for the PSS-I, $r_{xx} = .80$ (Foa et al., 1993). An RC larger than 1.96 reflects statistically reliable change. Based on the two-step criterion, individuals were classified as recovered (passed both Cut Score A and RC criterion), improved (passed RC criterion but not Cut Score A), unchanged (passed neither criterion), or deteriorated (passed RC criterion but symptom scores increased) for each follow-up interval. Agreement between the PCL-5 and PSS-I in categorizing individuals was calculated in terms of proportions and kappas. Kappa values of .21 to .40 indicate fair; .41 to .60, moderate; and .61 to .80, substantial agreement (Kraemer, Periyakoil, & Noda, 2002).

We conducted two additional sets of mixed effects analyses (SAS PROC MIXED) as alternative tests of the relationships between scores on the two measures. A detailed description and the results of these analyses are available in the online supplemental materials.

Diagnostic utility. Using Kraemer's (1987, 1992) signal detection methodology, we generated optimally efficient cut scores for the PCL-5 relative to PTSD diagnosis according to *DSM–IV* and *DSM–5* criteria, using various measures and scoring rules. In addition to using the PSS-I-based PTSD caseness rules described in the Method section, we selected cut scores of 39 (Dickstein et al., 2015), 44, and 50 (McDonald & Calhoun, 2010) to indicate a positive screen for PTSD on the PCL-S. We also generated optimally efficient cut scores for the PCL-S relative to the PCL-5 *DSM–5*-based (moderate or above) caseness rule. Optimally efficient scores maximize the number of agreements between test and diagnosis (i.e., minimize the number of false negatives and false positives).

Results

Descriptive Characteristics of the Sample

Descriptive characteristics of the sample are reported in Table 1. The mean age was 33 years, and the sample was diverse in race and military role. Most participants were male, were married, had completed some college, and were active duty enlisted personnel. Nearly all had deployed at least once to Iraq or Afghanistan, and they had, on average, last returned from deployment 2 years prior to completing the baseline assessment.

PTSD severity varied somewhat according to demographics and warzone exposure. Women had higher scores than men on the PCL-S ($M = 56.45$, $SD = 11.86$; $M = 52.80$, $SD = 12.27$), $t(910) = -2.51$, $p = .012$, and PSS-I ($M = 21.27$, $SD = 9.27$; $M = 19.32$, $SD = 9.53$), $t(910) = -2.71$, $p = .007$, at baseline. Older age was associated with higher scores on the PCL-5, $r = .12$, $p = .011$, and PCL-S, $r = .090$, $p = .047$, at follow-up. Participants who had deployed once ($M = 43.97$, $SD = 14.99$) or twice ($M = 43.82$, $SD = 14.74$) had higher baseline PCL-5 scores than those who deployed three ($M = 40.12$, $SD = 15.28$) or more than three ($M = 39.78$, $SD = 14.98$) times, $F(3, 906) = 4.85$, $p = .002$, and participants who deployed twice ($M = 54.45$, $SD = 12.13$) had higher PCL-S scores than those who deployed three times ($M = 51.37$, $SD = 11.91$), $F(3, 906) = 3.66$, $p = .012$. Finally, participants whose military occupational specialty duty was combat arms (e.g., infantry) had higher PSS-I scores at follow up ($M = 20.95$, $SD = 9.51$) than those who were in combat support (e.g., security; $M = 18.61$, $SD = 9.98$) or combat service support (e.g., mechanic; $M = 18.49$, $SD = 9.12$) roles, $F(2, 468) = 3.71$, $p = .025$.

Base rates of PTSD according to various definitions (see the Method section) are reported in Table 3 and ranged from 51% using the stringent PSS-I scoring rule to 84% using the lenient PSS-I scoring rule. High rates were expected because of the treatment-seeking nature of the sample. The agreement between various scoring rules for PTSD caseness and a positive diagnosis on the PCL-5 using the *DSM–5* scoring rule is reported as kappa values in Table 3. Agreement with the PCL-5 was highest for the PCL-S, moderate for the PSS-I using the lenient scoring rule, and lowest for the PSS-I using the stringent scoring rule.

Table 3
PTSD Prevalence in the Sample

Scoring rule for diagnosis	% (*n*)		Agreement with PCL-5 *DSM-5* diagnosis (κ)	
	Baseline	Follow-up	Baseline	Follow-up
PCL-5 *DSM-5* rule	74.1 (676)	24.1 (220)	N/A	N/A
PCL-S *DSM-IV* rule	71.7 (654)	25.4 (232)	.60	.71
PSS-I *DSM-IV* lenient rule	83.9 (765)	31.0 (283)	.43	.55
PSS-I *DSM-IV* stringent rule	51.2 (467)	17.4 (159)	.36	.45

Note. See Method section for definitions of scoring rules. PCL-5 = PTSD Checklist for *DSM-5*; *DSM-5* = fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); PCL-S = PTSD Checklist specific stressor version for *DSM-IV*; *DSM-IV* = fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994); PSS-I = PTSD Symptom Scale–Interview version; N/A = Not applicable.

Descriptive Characteristics of PCL-5 Scores

We report means and standard deviations for PCL-5 scores and all other study variables in Table 2. PCL-5 scores were distributed normally (skew statistic [standard error] = −.27 [.08] at baseline and .05 [.12] at follow-up).

Internal Consistency

As shown in Table 2, the PCL-5 overall severity and subscale scores demonstrated high internal consistency at baseline and follow-up, with Cronbach's alpha values ranging from .75 for the AR subscale at baseline to .95 for the overall scale at follow-up. Interitem correlations were analyzed as an additional index of internal consistency. In general, these fell in the recommended range of .15 to .50 (Clark & Watson, 1995), with a range of .10 to .74 ($M = .33$).

Validity

Convergent and discriminant validity. We present predicted and observed correlations between PCL-5 scores and scores on criterion measures in Table 4. The values of the effect size statistics $r_{\text{alerting-CV}}$ and $r_{\text{contrast-CV}}$ were .94 and .92, respectively, which indicate that the observed correlations strongly matched the predicted pattern of correlations.

Structural validity. Initial CFA results for the four-factor *DSM–5* model indicated poor fit. Modification indices indicated that correlating Item 1 with Item 2, Item10 with Item 11, and Item 17 with Item 18 would improve model fit. For the five-factor dysphoria and the six-factor externalizing behaviors models, modification indices indicated that correlating Item 1 with Item 2 and Item 10 with Item 11 would improve model fit. For the six-factor anhedonia and the seven-factor hybrid models, modification indices indicated that correlating Items 1 and 2 would improve model fit. These CFA results are shown in Table 5.

Comparing the nested models, the six-factor anhedonia and six-factor externalizing behaviors models fit the data better than the four-factor *DSM–5* model, $\Delta\chi^2(7) = 234.6$, $p < .001$, $\Delta\chi^2(8) = 45.76$, $p < .001$, and the five-factor dysphoric arousal model, $\Delta\chi^2(4) = 201.28$, $p < .001$, $\Delta\chi^2(5) = 12.44$, $p = .029$, respectively. Comparing the two six-factor models, which are non-nested, the anhedonia model more closely fit the data than the externalizing behaviors model, as evidenced by a 190.84-point AIC difference and 195.65-point BIC difference. The seven-factor hybrid model fit the data more closely than all the other models: the four-factor *DSM–5* model, $\Delta\chi^2(13) = 256.99$, $p < .001$, the five-factor dysphoric arousal model, $\Delta\chi^2(10) = 223.67$, $p < .001$, the six-factor externalizing behaviors model, $\Delta\chi^2(5) = 211.23$, $p < .001$, and the six-factor anhedonia model, $\Delta\chi^2(6) = 22.39$, $p = .001$.

Sensitivity to clinical change. Results in Table 6 show agreement between the PCL-5 and PSS-I in identifying pre–post changes of various magnitudes (i.e., 5- to 20-point improvements). Kappas ranged from .28 to .55, and the percentage of cases in agreement ranged from 72% for the 5-point change to 82% for the

Table 4
Predicted and Observed Correlations Between PCL-5 Total Severity Scores and Criterion Measures at Baseline

Variable	Observed *r*	Predicted *r*
PCL-S	.87*	.90
PSS-I	.68*	.80
DRRI Combat experiences	.12*	.30
DRRI Aftermath of battle	.22*	.30
BAI	.61*	.50
BDI	.64*	.60
TRGI		
Hindsight-bias/Responsibility	.32*	.30
Wrongdoing	.34*	.35
Lack of justification	.03	.02
STAXI-2 S-Anger	.33*	.35
ISI	.48*	.50
AUDIT	.10*	.25
PHQ-15	.49*	.50
RSES	–.22*	–.30

Note. See Table 2 for baseline *n*. PCL-5 = PTSD Checklist for *DSM-5;* PCL-S = PTSD Checklist specific stressor version for *DSM-IV*; PSS-I = PTSD Symptom Scale–Interview version; DRRI = Deployment Risk and Resilience Inventory; BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; TRGI = Trauma-Related Guilt Inventory; STAXI-2 S-Anger = State-Trait Anger Expression Inventory-2, State Anger subscale; ISI = Insomnia Severity Index; AUDIT = Alcohol Use Disorders Identification Test; PHQ-15 = Patient Health Questionnaire-15; RSES = Response to Stressful Experiences Scale.
* $p < .01$, 2-tailed.

15-point change. The Pearson correlation coefficient between the PCL-5 and PSS-I pre–post change scores was .72, and the concordance correlation coefficient was .68, indicating that the change scores were highly correlated and of nearly equal magnitude.

Results in Table 7 show agreement between PCL-5 and PSS-I in identifying statistically reliable and clinically significant change; shown are the numbers and percentages of individuals for whom categorizations overlapped, as well as the numbers of individuals classified in each category. The cut score indicating clinically meaningful improvement on the PCL-5 was 24, and the cut score indicating clinically significant exacerbation was 76. The respective cut scores for the PSS-I were 13 and 42; values used in the analyses based on rescaling the PSS-I to match the 5-point scale of the PCL-5 were 18 and 56. Using these cut scores and the RC criteria to classify individuals into categories of change, the PCL-5 and PSS-I classifications matched for 65% of individuals (simple $\kappa = .41$, 95% confidence interval [CI] [.27, .56]; weighted $\kappa = .51$, 95% CI [.32, .71]), indicating moderate agreement between the measures.

Diagnostic Utility

In Table 8, we provide optimally efficient cut scores on the PCL-5 relative to PTSD diagnosis according to *DSM–IV* and *DSM–5* criteria, using various measures and scoring rules. A PCL-5 score of 33 and a PCL-S score of 43 were found to be optimally efficient for detecting PTSD cases according to PCL-5 *DSM–5* scoring criteria. PCL-5 scores of 36, 23, and 42 were found to be optimally efficient for detecting PTSD cases according to PCL-S *DSM–IV* scoring, PSS-I lenient scoring, and PSS-I stringent scoring criteria, respectively. PCL-5 scores of 25, 31, and 39 were found to be optimally efficient in corresponding to PCL-S cut scores of 39, 44, and 50, respectively.

Discussion

As criteria for PTSD have evolved, the PCL required revision to be consistent with *DSM–5*. We examined the psychometric properties of the PCL-5 in a large cohort ($N = 912$) of service members seeking treatment for PTSD while stationed in garrison (i.e., their permanent duty station). The PCL-5 scores were found to be psychometrically sound when evaluated in a clinical setting with high base rates for PTSD. The PCL-5 total and subscale scores were found to have high internal consistency. In terms of construct validity, effect size estimates indicated that predicted correlations, based on published research with the PCL-S, strongly matched observed correlations between the PCL-5 and criterion measures. As expected, the largest positive correlation emerged between PCL-5 scores and PCL-S scores, as they share most items (see Hoge et al., 2014) as well as method variance. The next largest correlation was with the clinician-administered measure of PTSD, suggesting strong convergent validity. Correlations with measures of other constructs were weaker and followed the predicted pattern. Shared method (i.e., self-report) variance may have inflated correlations.

The correlations between scores on the PCL-5 and scores on the Combat Experiences and Aftermath of Battle subscales of the DRRI were smaller than predicted. The predicted correlations (rs = .30) were based on the DRRI validation study, which used a prior version of the PCL that did not require symptoms to be

Table 5
Fit Statistics for PCL-5 Factor Models at Baseline

Model	$\chi^2(df)$	RMSEA	CFI	TLI	SRMR	AIC	BIC	$\Delta\chi^2(df)$	*p*
DSM-5	861.70 (161)	.069	.908	.892	.059	52,833.17	53,165.45		
Five-factor dysphoric arousal	828.38 (158)	.068	.912	.894	.057	52,805.84	53,152.57	33.32 (3)	<.001
Six-factor externalizing behaviors	815.94 (153)	.069	.913	.892	.056	52,803.41	53,174.21	12.44 (5)	.029
Six-factor anhedonia	627.10 (154)	.058	.938	.924	.042	52,612.57	52,978.56	201.28 (4)	<.001
Seven-factor hybrid	604.71 (148)	.058	.940	.923	.041	52,602.17	52,997.06	211.23 (5)	<.001

Note. Chi-square difference results represent the following comparisons: the five-factor model compared with the four-factor model, both six-factor models compared with the five-factor model, and the seven-factor model compared with the six-factor anhedonia model. PCL-5 = PTSD Checklist for *DSM-5*; *DSM*-5 = fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); *df* = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

Table 6
Pairwise Agreement Between PCL-5 and PSS-I in Identifying Change Using Arbitrary Cut Points

Cut point	κ	*SE*	95% CI	% of cases in agreement
5 points	.45	.093	[.27, .63]	72.53
10 points	.53	.092	[.34, .71]	78.02
15 points	.55	.097	[.36, .74]	82.42
20 points	.28	.108	[.07, .49]	78.02

Note. Percentage of cases in agreement are those for which changes on the PCL-5 and PSS-I are either both above or both below the cut point. Discordant pairs have one change score above and the other below the cut point. The kappa statistic, κ, has standard error, *SE*. PCL-5 = PTSD Checklist for *DSM-5*; PSS-I = PTSD Symptom Scale–Interview version; CI = confidence interval.

indexed to a specific Criterion A index event (Vogt et al., 2008). The strength of the relationship between the DRRI exposure subscales and PCL severity has been considerably smaller in studies that have asked service members to index their PTSD symptoms to a single Criterion A event (e.g., Nash et al., 2015). When, as in the present study, all participants have had sufficient exposure to warrant a PTSD interview, and they are furthermore required to index their PTSD symptoms to a single event, the expectation of a moderate dose–response relationship between broad exposure scales and PTSD symptom severity should be reconsidered.

Consistent with existing evaluations of the PCL-5 factor structure, the *DSM–5* model had only adequate fit (Kline, 2005). In fact, no previously evaluated alternative factor structure was found to have close fit in this study. Nevertheless, a seven-factor model showed superior fit to the *DSM–5* model and other variations, which is consistent with Armour et al. (2015). It may be necessary to reconsider the *DSM–5* symptom factors (clusters) if they consistently fail to be empirically validated. Alternatively, it could be difficult to replicate the putative subcomponents of *DSM–5*-defined PTSD with a self-report measure; *DSM–5*-based structured clinical interviews may fare better.

The PCL-5 was comparable with an interview-based assessment of PTSD symptoms in terms of sensitivity to clinical change (pre- to posttreatment). PCL-5 and PSS-I scores showed a fair to moderate correspondence in categorizing change based on arbitrary cut points and high concordance between pre–post change scores. Using the conservative Jacobson and Truax (1991) method to categorize individuals based on reliable and clinically significant change, there was moderate agreement between the PCL-5 and PSS-I. The measures agreed most in classifying individuals who did not change or showed clinically significant improvement. The PCL-5's classification of a smaller number of individuals into the clinically significant improvement category suggests that it may be more conservative than the PSS-I in identifying clinically meaningful change.

A thorough examination of substantive and clinically meaningful change on the PCL-5 was beyond the scope of this effort. That would require clinician or patient judgments and referencing scores to healthy outcomes (Jacobson, Roberts, Berns, & McGlinchey, 1999; Jacobson & Truax, 1991; Kraemer et al., 2003). Other researchers have proposed simpler methods for indexing clinically significant change (e.g., 0.5 *SD*; Norman, Sloan, & Wyrwich, 2003). Monson et al. (2008) reported clinically significant change on a *DSM–IV* version of PCL to be 10 points, stating that they used the Jacobson and Truax (1991) method. However, it is unclear what these authors used to index healthy scores on the PCL, and they did not specify the formulas used to derive this result, given that the Jacobson and Truax method results in a cut score as opposed to a change-score. When the range of scores typical of the healthy population is unknown, Jacobson and Truax recommend a conservative strategy to estimate clinically meaningful change, which entails calculating a cut score that is two standard deviations below mean baseline scores for patients entering a clinical trial. In this study, this approach yielded a score of 24; for service members getting treatment for PTSD while in garrison, posttreatment scores at or below 24 on the PCL-5 likely represent clinically significant change, provided their RC index also exceeds 1.96 (see Jacobson & Truax, 1991). We provide this information with a note of caution. It is difficult to draw conclusions from this variation of the Jacobson and Truax method because it does not adjust for dropout, may capitalize on regression to the mean, and cannot be used with non-normal distributions of scores. Additional limitations of this approach are reviewed elsewhere (e.g., Wise, 2004). Jacobson and colleagues (1999) also recommend that clinically significant cut scores be standardized by aggregating results from various studies, which awaits further research.

Signal detection test results suggested a variety of optimally efficient cut scores on the PCL-5 for predicting PTSD diagnosis according to *DSM–IV* and *DSM–5* criteria, using various measures

Table 7
Agreement Between PCL-5 and PSS-I in Identifying Reliable and Clinically Significant Change (n [%])

Clinical change category	1	2	3	4	5	Total *n* for PSS-I
1. Clinically significant improvement	11 (84.6%)					21
2. Statistically reliable improvement		6 (33.3%)				12
3. No change			41 (78.8%)			54
4. Statistically reliable exacerbation				1 (16.7%)		3
5. Clinically significant exacerbation					0 (0%)	1
Total *n* for PCL-5	13	18	52	6	2	91

Note. Numbers and percentages on the diagonal represent the degree to which PCL-5 categorization agreed with PSS-I categorization, equating to 59 or 65% of pairs overall. Simple kappa = .41, 95% CI [.27, .56]. Weighted kappa = .51, 95% CI [.32, .71]. PCL-5 = PTSD Checklist for *DSM-5*; *DSM-5* = fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); PSS-I = PTSD Symptom Scale–Interview version; CI = confidence interval.

Table 8
Diagnostic Utility of Optimally Efficient Cut Scores on the PTSD Checklist (PCL-5 and PCL-S) for Predicting PTSD Diagnosis Based on Various Criteria

Cut score	Level (%)	Sensitivity	Specificity	PPV	NPV	Eff	κ(0)	κ (.5)	95% CI	κ (1)	AUC
				PCL-5 *DSM-5* rule (base rate = 74.12%)							
PCL-5 total											
33	76.32	.93	.72	.91	.79	.88	.64	.68	[.62, .73]	.72	.930
				PCL-S *DSM-IV* rule (base rate = 71.71%)							
36	68.42	.86	.76	.90	.68	.83	.64	.59	[.54, .65]	.55	.887
				PSS-I lenient rule (base rate = 83.88%)							
23	87.72	.95	.48	.90	.63	.87	.40	.47	[.39, .55]	.55	.836
				PSS-I stringent rule (base rate = 51.21%)							
42	54.93	.77	.68	.72	.74	.73	.42	.45	[.39, .51]	.49	.800
				PCL-S score of 39 (base rate = 87.28%)							
25	86.51	.95	.75	.96	.71	.93	.71	.69	[.62, .76]	.66	.955
				PCL-S score of 44 (base rate = 78.18%)							
31	79.28	.95	.77	.94	.81	.91	.71	.74	[.68, .79]	.76	.947
				PCL-S score of 50 (base rate = 62.17%)							
39	61.29	.89	.85	.91	.83	.88	.75	.74	[.69, .78]	.72	.942
				PCL-5 *DSM-5* rule (base rate = 74.12%)							
PCL-S total											
43	80.04	.94	.60	.87	.77	.85	.50	.58	[.52, .64]	.70	.872

Note. $N = 912$. Values rounded to decimal places shown. Confidence intervals provided for κ (.5). Measures of test quality are adjusted for chance agreement between the test and criterion. These values range from .00 (chance agreement) to 1.00 (perfect agreement). PCL-5 = PTSD Checklist for *DSM-5*; PCL-S = PTSD Checklist specific stressor version for *DSM-IV*; PTSD = posttraumatic stress disorder; Level = level of test (i.e., % of participants meeting cutoff); PPV = positive predictive value; NPV = negative predictive value; Eff = efficiency; κ (0) = quality of specificity; κ (.5) = quality of efficiency; κ (1) = quality of sensitivity; CI = confidence interval; AUC = area under the curve; *DSM-5* = fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); *DSM*-IV = fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994); PSS-I = PTSD Symptom Scale–Interview version.

and scoring rules. The optimally efficient cut scores identified in this study were consistent with other research. For example, Hoge et al. (2014) found a score of 32 on the PCL-5 had optimum agreement with a PCL-S score of 44, and a score of 38 on the PCL-5 had optimum agreement with a PCL-S score of 50. Analysis of kappa (.5) values, which index quality of efficiency, suggest that correspondences between the PCL-5 and the PSS-I-based diagnoses are lowest, because of the instruments' different modalities (i.e., self-report vs. clinician-administered) and different diagnostic criteria (i.e., *DSM–5* vs. *DSM–IV*). The correspondence of the PCL-5 cut score with the PCL-S *DSM–IV* diagnosis is higher, and higher still with the PCL-5 *DSM–5* diagnosis. Finally, the correspondence of the PCL-5 cut score with specific PCL-S cut score criteria is the highest of all, and in the excellent range for kappa. That the correspondence of the PCL-S cut score with PCL-5 *DSM–5* diagnosis is essentially equivalent to the correspondence of the PCL-5 cut score with the PCL-S *DSM–IV* diagnosis provides evidence of the strength of association between *DSM–IV* and *DSM–5* criteria.

That we could not evaluate order effects of the presentation of the PCL versions is a significant limitation of this study. Ideally, the PCL-S and PCL-5 presentation order would have been counterbalanced (as in Hoge et al., 2014), with both measures presented prior to the PSS-I to minimize carryover effects from the interview to the self-report measures. We also could not assess the test–retest reliability of the PCL-5, because the PCL-5 was administered once at baseline and again after treatment. Although there is no reason to suspect poor performance in this area (e.g., the PCL-S showed adequate test–retest reliability; Wilkins et al., 2011), future studies should assess this characteristic of the PCL-5.

In summary, we found the PCL-5 to have very sound psychometric properties overall, although, consistent with prior research, the factor structure of the scale did not conform optimally to *any* existing proposed factor structure. The results of this psychometric analysis of PCL-5 scores are not generalizable to contexts in which the base rates for PTSD are lower than our study, and the properties of the PCL-5 scores, particularly cut scores for screening and diagnosis, are also likely to vary by sample and context (see McDonald & Calhoun, 2010). Consequently, our results are chiefly applicable to clinical settings that treat active duty military service members and new veterans seeking care for PTSD.

References

Afari, N., Pittman, J., Floto, E., Owen, L., Buttner, M., Hossain, N., . . . Lohr, J. B. (2015). Differential impact of combat on postdeployment symptoms in female and male veterans of Iraq and Afghanistan. *Military*

Medicine, 180, 296–303. http://dx.doi.org/10.7205/MILMED-D-14-00255

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

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

Armour, C., Tsai, J., Durham, T. A., Charak, R., Biehn, T. L., Elhai, J. D., & Pietrzak, R. H. (2015). Dimensional structure of DSM–5 posttraumatic stress symptoms: Support for a hybrid Anhedonia and Externalizing Behaviors model. *Journal of Psychiatric Research, 61,* 106–113. http://dx.doi.org/10.1016/j.jpsychires.2014.10.012

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The Alcohol Use Disorders Identification Test: Guidelines for use in primary care* (2nd ed.). Geneva, Switzerland: World Health Organization.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56,* 893–897. http://dx.doi.org/10.1037/0022-006X.56.6.893

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the BDI-II*. San Antonio, TX: Psychological Corporation.

Biehn, T. L., Elhai, J. D., Seligman, L. D., Tamburrino, M., Armour, C., & Forbes, D. (2013). Underlying dimensions of DSM–5 posttraumatic stress disorder and major depressive disorder symptoms. *Psychological Injury and Law, 6,* 290–298. http://dx.doi.org/10.1007/s12207-013-9177-4

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,* 75–90. http://dx.doi.org/10.1002/jts.2490080106

Bovin, M., Marx, B., Weathers, F., Gallagher, M., Rodriguez, P., Schnurr, P., & Keane, T. (in press). Psychometric properties of the PTSD Checklist for *DSM-5* (PCL-5) in veterans. *Psychological Assessment*.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33,* 261–304. http://dx.doi.org/10.1177/0049124104268644

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

Crawford, E. F., Fulton, J. J., Swinkels, C. M., Beckham, J. C., Calhoun, P. S., & VA Mid-Atlantic MIRECC OEF/OIF Registry Workgroup. (2013). Diagnostic efficiency of the AUDIT-C in U.S. veterans with military service since September 11, 2001. *Drug and Alcohol Dependence, 132,* 101–106. http://dx.doi.org/10.1016/j.drugalcdep.2013.01.012

Dickstein, B. D., Weathers, F. W., Angkaw, A. C., Nievergelt, C. M., Yurgil, K., Nash, W. P., . . . Marine Resiliency Study Team. (2015). Diagnostic utility of the Posttraumatic Stress Disorder (PTSD) Checklist for identifying full and partial PTSD in active-duty military. *Assessment, 22,* 289–297. http://dx.doi.org/10.1177/1073191114548683

Elhai, J. D., Biehn, T. L., Armour, C., Klopper, J. J., Frueh, B. C., & Palmieri, P. A. (2011). Evidence for a unique PTSD construct represented by PTSD's D1–D3 symptoms. *Journal of Anxiety Disorders, 25,* 340–345. http://dx.doi.org/10.1016/j.janxdis.2010.10.007

Epstein, D., Goren, K. J., & Bushnell, M. (2009). Preliminary evaluation of the Insomnia Severity Index in Operation Enduring Freedom/Operation Iraqi Freedom veterans. *Sleep, 32,* A272.

Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress, 6,* 459–473. http://dx.doi.org/10.1002/jts.2490060405

Foa, E. B., & Tolin, D. F. (2000). Comparison of the PTSD Symptom Scale-Interview Version and the Clinician-Administered PTSD scale. *Journal of Traumatic Stress, 13,* 181–191. http://dx.doi.org/10.1023/A:1007781909213

Gray, M. J., Litz, B. T., Hsu, J. L., & Lombardo, T. W. (2004). Psychometric properties of the life events checklist. *Assessment, 11,* 330–341. http://dx.doi.org/10.1177/1073191104269954

Hoge, C. W., Riviere, L. A., Wilk, J. E., Herrell, R. K., & Weathers, F. W. (2014). The prevalence of post-traumatic stress disorder (PTSD) in US combat soldiers: A head-to-head comparison of DSM–5 versus DSM–IV–TR symptom criteria with the PTSD checklist. *The Lancet Psychiatry, 1,* 269–277. http://dx.doi.org/10.1016/S2215-0366(14)70235-4

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology, 67,* 300–307. http://dx.doi.org/10.1037/0022-006X.67.3.300

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19. http://dx.doi.org/10.1037/0022-006X.59.1.12

Johnson, D. C., Polusny, M. A., Erbes, C. R., King, D., King, L., Litz, B. T., . . . Southwick, S. M. (2011). Development and initial validation of the Response to Stressful Experiences Scale. *Military Medicine, 176,* 161–169. http://dx.doi.org/10.7205/MILMED-D-10-00258

Keane, T. M., Rubin, A., Lachowicz, M., Brief, D., Enggasser, J. L., Roy, M., . . . Rosenbloom, D. (2014). Temporal stability of DSM–5 posttraumatic stress disorder criteria in a problem-drinking sample. *Psychological Assessment, 26,* 1138–1145. http://dx.doi.org/10.1037/a0037133

King, L. A., King, D. W., Vogt, D. S., Knight, J., & Samper, R. E. (2006). Deployment Risk and Resilience Inventory: A collection of measures for studying deployment-related experiences of military personnel and veterans. *Military Psychology, 18,* 89–120. http://dx.doi.org/10.1207/s15327876mp1802_1

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Kraemer, H. C. (1987). The methodological and statistical evaluation of medical tests: The dexamethasone suppression test in psychiatry. *Psychoneuroendocrinology, 12,* 411–427. http://dx.doi.org/10.1016/0306-4530(87)90076-X

Kraemer, H. C. (1992). *Evaluating medical tests: Objective and quantitative guidelines*. Newbury Park, CA: Sage.

Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *Journal of the American Academy of Child & Adolescent Psychiatry, 42,* 1524–1529. http://dx.doi.org/10.1097/00004583-200312000-00022

Kraemer, H. C., Periyakoil, V. S., & Noda, A. (2002). Kappa coefficients in medical research. *Statistics in Medicine, 21,* 2109–2129. http://dx.doi.org/10.1002/sim.1180

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine, 64,* 258–266. http://dx.doi.org/10.1097/00006842-200203000-00008

Kubany, E. S., Haynes, S. N., Abueg, F. R., Manke, F. P., Brennan, J. M., & Stahura, C. (1996). Development and validation of the Trauma-Related Guilt Inventory (TRGI). *Psychological Assessment, 8,* 428–444. http://dx.doi.org/10.1037/1040-3590.8.4.428

Liu, P., Wang, L., Cao, C., Wang, R., Zhang, J., Zhang, B., . . . Elhai, J. D. (2014). The underlying dimensions of *DSM–5* posttraumatic stress disorder symptoms in an epidemiological sample of Chinese earthquake survivors. *Journal of Anxiety Disorders, 28,* 345–351.

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD checklist: A critical review. *Clinical Psychology Review, 30,* 976–987. http://dx.doi.org/10.1016/j.cpr.2010.06.012

Miller, M. W., Wolf, E. J., Kilpatrick, D., Resnick, H., Marx, B. P., Holowka, D. W., . . . Friedman, M. J. (2013). The prevalence and latent structure of proposed DSM–5 posttraumatic stress disorder symptoms in US national and veteran samples. *Psychological Trauma: Theory, Research, Practice, and Policy, 5,* 501–512. http://dx.doi.org/10.1037/a0029730

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,* 131–138. http://dx.doi.org/10.1037/1040-3590.20.2.131

Morin, C. M. (1993). *Insomnia: Psychological assessment and management*. New York, NY: Guilford Press.

Muthén, L. K., & Muthén, B. O. (2013). *Mplus (Version 7.11)* [Computer software]. Los Angeles, CA: Author.

Nash, W. P., Boasso, A. M., Steenkamp, M. M., Larson, J. L., Lubin, R. E., & Litz, B. T. (2015). Posttraumatic stress in deployed Marines: Prospective trajectories of early adaptation. *Journal of Abnormal Psychology, 124,* 155–171. http://dx.doi.org/10.1037/abn0000020

Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care, 41,* 582–592. http://dx.doi.org/10.1097/01.MLR.0000062554.74615.4C

Pietrzak, R. H., Tsai, J., Armour, C., Mota, N., Harpaz-Rotem, I., & Southwick, S. M. (2015). Functional significance of a novel 7-factor model of DSM–5 PTSD symptoms: Results from the National Health and Resilience in Veterans study. *Journal of Affective Disorders, 174,* 522–526. http://dx.doi.org/10.1016/j.jad.2014.12.007

Reinert, D. F., & Allen, J. P. (2007). The alcohol use disorders identification test: An update of research findings. *Alcoholism: Clinical and Experimental Research, 31,* 185–199. http://dx.doi.org/10.1111/j.1530-0277.2006.00295.x

Spielberger, C. D. (1999). *State-Trait Anger Expression Inventory-2 (STAXI-2): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Taft, C. T., Street, A. E., Marshall, A. D., Dowdall, D. J., & Riggs, D. S. (2007). Posttraumatic stress disorder, anger, and partner abuse among Vietnam combat veterans. *Journal of Family Psychology, 21,* 270–277. http://dx.doi.org/10.1037/0893-3200.21.2.270

Tsai, J., Harpaz-Rotem, I., Armour, C., Southwick, S. M., Krystal, J. H., & Pietrzak, R. H. (2015). Dimensional structure of DSM–5 posttraumatic stress disorder symptoms: Results from the National Health and Resilience in Veterans Study. *Journal of Clinical Psychiatry, 76,* 546–553. http://dx.doi.org/10.4088/JCP.14m09091

Vasterling, J. J., Proctor, S. P., Friedman, M. J., Hoge, C. W., Heeren, T., King, L. A., & King, D. W. (2010). PTSD symptom increases in Iraq-deployed soldiers: Comparison with nondeployed soldiers and associations with baseline symptoms, deployment experiences, and postdeployment stress. *Journal of Traumatic Stress, 23,* 41–51.

Vogt, D. S., Proctor, S. P., King, D. W., King, L. A., & Vasterling, J. J. (2008). Validation of scales from the Deployment Risk and Resilience Inventory in a sample of Operation Iraqi Freedom veterans. *Assessment, 15,* 391–403. http://dx.doi.org/10.1177/1073191108316030

Vogt, D., Smith, B. N., King, D. W., & King, L. A. (2012). *Manual for the Deployment Risk and Resilience Inventory-2 (DRRI-2): A collection of measures for studying deployment-related experiences of military veterans*. Boston, MA: National Center for PTSD.

Weathers, F., Litz, B., Herman, D., Huska, J., & Keane, T. (1993, October). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility*. Paper presented at the annual meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

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). Retrieved from http://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp

Weathers, F. W., Marx, B. P., Friedman, M. J., & Schnurr, P. P. (2014). Posttraumatic stress disorder in DSM–5: New criteria, new measures, and implications for assessment. *Psychological Injury and Law, 7,* 93–107. http://dx.doi.org/10.1007/s12207-014-9191-1

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

Wilkins, K. C., Lang, A. J., & Norman, S. B. (2011). Synthesis of the psychometric properties of the PTSD checklist (PCL) military, civilian, and specific versions. *Depression and Anxiety, 28,* 596–606. http://dx.doi.org/10.1002/da.20837

Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment, 82,* 50–59. http://dx.doi.org/10.1207/s15327752jpa8201_10

Received July 27, 2015
Revision received October 21, 2015
Accepted October 21, 2015 ■