Psychological Assessment
1999, Vol. 11, No. 2, 124–133

In the public domain

Psychometric Properties of Nine Scoring Rules for the Clinician-Administered Posttraumatic Stress Disorder Scale

Frank W. Weathers
Auburn University

Ayelet Meron Ruscio
Boston Veterans Affairs Medical Center

Terence M. Keane
Boston Veterans Affairs Medical Center and Boston University

The use of structured interviews that yield continuous measures of symptom severity has become increasingly widespread in the assessment of posttraumatic stress disorder (PTSD). To date, however, few scoring rules have been developed for converting continuous severity scores into dichotomous PTSD diagnoses. In this article, we describe and evaluate 9 such rules for the Clinician-Administered PTSD Scale (CAPS). Overall, these rules demonstrated good to excellent reliability and good correspondence with a PTSD diagnosis based on the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev.; *DSM–III–R*; American Psychiatric Association, 1987). However, the rules yielded widely varying prevalence estimates in 2 samples of male Vietnam veterans. Also, the use of *DSM–III–R* versus *DSM–IV* criteria had negligible impact on PTSD diagnostic status. The selection of CAPS scoring rules for different assessment tasks is discussed.

A growing trend in the assessment of posttraumatic stress disorder (PTSD) is the use of structured interviews that use dimensional rather than categorical (present or absent) rating scales to evaluate PTSD symptom severity. Examples of such interviews include the Structured Interview for PTSD (SI-PTSD; Davidson, Smith, & Kudler, 1989), the PTSD Symptom Scale Interview (PSS-I; Foa, Riggs, Dancu, & Rothbaum, 1993), and the Clinician-Administered PTSD Scale (CAPS; Blake et al., 1990, 1995). An advantage of these interviews over instruments such as the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; SCID; First, Spitzer, Gibbon, & Williams, 1997) is that they yield continuous measures of PTSD symptom severity—for individual symptoms, symptom clusters, and the entire syndrome—as well as a dichotomous PTSD diagnosis.[1] By assessing finer gradations of symptom severity, these interviews can differentiate individuals with incapacitating symptoms from those who just exceed the diagnostic threshold, and they can differentiate individuals with subthreshold but clinically significant symptoms from those who are essentially asymptomatic. Dimensional interviews also make it possible to track subtle changes in symptom severity over time, which is crucial for treatment outcome studies and other longitudinal research designs. Finally, such measures offer greater flexibility for statistical analyses: Continuous severity scores permit the computation of means and provide greater variability for correlational analyses, multiple regression analyses, and factor analyses.

Despite the advantages of continuous measures of PTSD symptom severity, a number of clinical and research assessment tasks call for a dichotomous PTSD diagnosis (for a discussion of categorical vs. dimensional approaches in the assessment of psychopathology, see Blashfield, 1984; Lorr, 1986; Widiger, 1997). In clinical assessments, a PTSD diagnosis is used to summarize and conceptualize individual symptoms, select and implement appropriate interventions, communicate with other clinicians, and provide documentation to insurance companies and health maintenance organizations. In epidemiological research, a diagnosis is used to estimate the prevalence of PTSD; in case-control research it is used to create relatively homogeneous comparison groups. In

Frank W. Weathers, Department of Psychology, Auburn University; Ayelet Meron Ruscio, National Center for Posttraumatic Stress Disorder (PTSD)—Behavioral Science Division, Boston Veterans Affairs Medical Center; Terence M. Keane, National Center for PTSD—Behavioral Science Division, Boston Veterans Affairs Medical Center, and School of Medicine, Boston University.

Ayelet Meron Ruscio is now at Department of Psychology, Pennsylvania State University.

Portions of this article were presented at the annual meeting of the International Society for Traumatic Stress Studies, Montreal, Quebec, Canada, November 1997.

Correspondence concerning this article should be addressed to Frank W. Weathers, Department of Psychology, 226 Thach Hall, Auburn University, Alabama 36849-5214. Electronic mail may be sent to weathfw@mail.auburn.edu.

[1]Although the SCID is a diagnostic instrument, intended primarily for assessing the presence or absence of psychiatric disorders, the SCID PTSD module can be used to create a continuous measure of PTSD severity by summing over the 17 items, as one of the reviewers noted. However, we are not aware of any studies that have empirically validated the SCID PTSD module for this purpose. Further, although this use of the SCID might be effective at the syndrome level, or possibly even at the symptom cluster level, the SCID does not provide a continuous severity measure for individual PTSD symptoms.

these and similar applications, there is a need to designate individuals as either PTSD positive (case) or PTSD negative (noncase or control). Therefore, when dimensional interviews are used in these contexts, the continuous severity scores they yield must be converted into a dichotomous diagnosis. On the CAPS, the complexity of this conversion is compounded by the fact that PTSD symptoms are rated on two separate dimensions of symptom severity: frequency and intensity.

A key question largely ignored by clinical investigators is how best to accomplish the necessary conversion from continuous scores to a dichotomous diagnosis. One approach is to dichotomize severity scores at the item level, creating a present or absent rating for each PTSD symptom, then follow the *DSM–IV* diagnostic algorithm (one reexperiencing symptom, three avoidance and numbing symptoms, and two hyperarousal symptoms) to obtain a diagnosis. A second approach is to sum across all items to obtain a total severity score, then select a cutoff score indicative of a PTSD diagnosis. With either approach, the use of different scoring rules results in classifying different groups of individuals as having PTSD. This can lead to widely varying prevalence estimates and can also affect conclusions about the phenomenology of PTSD, because those identified as PTSD positive by different scoring rules may differ substantively in their clinical presentation.

For example, Blanchard et al. (1995) evaluated three scoring rules for the CAPS and found that prevalence estimates ranged from 27% for the most stringent rule to 44% for the most lenient. They also found that participants who met PTSD criteria according to the most stringent scoring rule reported greater subjective distress and functional impairment than those who met criteria by a more lenient rule. This suggests that those identified as PTSD positive by one scoring rule may differ in important ways from those identified as PTSD positive by a different rule.

A second consideration for dichotomizing continuous scores is that scoring rules may be derived either rationally or empirically. Rationally derived rules are based on expert judgment about what makes sense to use, and thus they require clinical experience and inspection of the rating-scale anchors. Empirically derived rules are based on a statistical correspondence of PTSD symptom severity scores with some well-established criterion. To date, investigators who have developed dimensional interviews typically have generated and evaluated a single rationally derived cutoff for individual items, in some cases adding a single empirically derived cutoff for total severity. For example, for the SI-PTSD, which uses a 5-point rating scale (0 = *absent*, 1 = *mild*, 2 = *moderate*, 3 = *severe*, and 4 = *extremely severe*), Davidson et al. (1989) proposed that a PTSD symptom be considered present when an item is rated as 2 (*moderate*) or higher. In addition, they proposed a cutoff in the range of 16–18—for the 13-item *DSM–III* version of the scale—for converting the total severity score into a PTSD diagnosis.

Similarly, for the PSS-I, which uses a 4-point scale for individual items (0 = *not at all*, 1 = *a little bit*, 2 = *somewhat*, and 3 = *very much*), Foa et al. (1993) proposed a cutoff of 1 (*a little bit*) or higher for individual items. They did not identify an optimal cutoff for total severity. On the CAPS, the frequency and intensity of each PTSD symptom are rated on separate 5-point scales ranging from 0 to 4. Blake et al. (1990) proposed that a symptom be considered present when an item is rated with a frequency of 1 (*once a month*) or higher and an intensity of 2 (*moderate*) or higher. Weathers et al. (1998) identified a total severity score of 65 as optimal for predicting a PTSD diagnosis.

These scoring rules seem reasonable and appear to perform well psychometrically, although more cross-validation is needed to determine their stability and generalizability across different trauma populations and settings. Nonetheless, because dimensional interviews provide much greater flexibility in quantifying PTSD symptom severity, numerous alternative rules could be developed, some of which might prove to have more robust psychometric properties than the original rules. Therefore, it is crucial to develop multiple scoring rules for a given instrument and compare their utility for different assessment tasks.

Kraemer (1992) identified three types of tests, each of which is optimal for a different assessment task. Optimally sensitive tests, which minimize false negatives, are best for screening. Optimally specific tests, which minimize false positives, are best for confirming a diagnosis. Optimally efficient tests, which minimize overall number of diagnostic errors, giving equal weight to false positives and false negatives, are best for differential diagnosis. To date, research on dimensional PTSD interviews has focused almost exclusively on optimally efficient tests and differential diagnosis. However, screening for PTSD and confirming a PTSD diagnosis are also valuable assessment tasks and deserve greater attention. It is unlikely that a single scoring rule for a dimensional measure would be optimal for all three assessment tasks, which means that multiple scoring rules are needed to serve a variety of functions.

Our primary purpose in this article was to describe nine different scoring rules for the CAPS and investigate their reliability, their utility for the three different assessment tasks, and their estimated prevalence of PTSD. We also sought to explore the impact of using *DSM–III–R* versus *DSM–IV* diagnostic criteria for PTSD. This is important for two reasons. First, the field is still in transition from *DSM–III–R* to *DSM–IV*, and although the *DSM–IV* revisions of the PTSD criteria were relatively minor, and thus could be expected to have little impact on diagnostic decision making, there is little empirical evidence bearing on their equivalence to the *DSM–III–R* criteria. Second, because data collection for this study extended over a 6-year period that included the transition to *DSM–IV*, some participants were assessed using *DSM–III–R* criteria and others were assessed using *DSM–IV* criteria. We wanted to use *DSM–IV* criteria for all participants if this could be justified empirically.

Method

Participants

Participants included five samples of male Vietnam theater veterans evaluated at the National Center for PTSD at the Boston Veterans Affairs Medical Center. Table 1 presents demographic information for all participants.[2] Sample 1 consisted of 123 veterans recruited for a research project on the psychometric properties of the CAPS (Weathers et al., 1998). As described in Weathers et al. (1998), all participants in Sample 1 were first administered the Structured Clinical Interview for *DSM–III–R* (SCID;

[2] In addition to Weathers et al. (1998), portions of the data from the participants in Sample 1 were included in Herman, Weathers, Litz, and Keane (1996), Orsillo et al. (1996), Weathers et al. (1996), and Litz et al. (1997). Portions of the data from the participants in Sample 5 were included in D. W. King, Leskin, King, and Weathers (1998).

Table 1
Demographic Characteristics of the Five Samples

Variable	Sample 1 (n = 123)	2 (n = 24)	3 (n = 53)	4 (n = 67)	5 (n = 571)
Age (years)					
M	43.74	50.71	49.51	50.98	47.33
SD	2.69	4.78	5.57	4.59	8.82
Ethnicity (%)					
Caucasian	74.4	75.0	84.9	84.1	82.6
Black	0.8	20.8	9.4	11.0	12.3
Hispanic	0.0	0.0	0.0	0.0	2.5
Native American/Alaskan	23.1	4.2	1.9	2.4	1.8
Other	1.7	0.0	3.8	2.4	0.8
Military branch (%)[a]					
Army	48.4	37.5	47.2	58.5	54.6
Marines	29.5	16.7	26.4	25.6	29.6
Navy	13.9	16.7	13.2	13.4	10.7
Air Force	7.4	29.2	13.2	7.3	7.5
Other	0.8	0.0	0.0	1.2	1.8
Employment, any current (%)	37.4	43.5	48.1	58.5	43.2
Education (%)					
< High school diploma	10.7	4.2	1.9	13.4	11.5
High school diploma/GED	24.8	4.1	13.2	9.0	18.7
Some college/vocational	49.6	50.0	84.9	43.3	54.9
BA/BS or more	14.9	41.7	0.0	34.3	14.9
Marital status (%)					
Single (never married)	26.2	20.8	18.9	11.0	17.9
Married/live with partner	28.7	45.8	54.7	59.8	48.4
Separated/divorced	41.8	33.3	26.4	29.3	32.1
Widowed/other	3.3	0.0	0.0	0.0	1.6

Note. GED = Graduate Equivalency Diploma.
[a] Percentages summing to over 100% reflect service in multiple military branches by several individuals.

Spitzer, Williams, Gibbon, & First, 1990) PTSD module, followed by the CAPS 2 to 3 days later, by independent clinicians. In addition, the first 60 participants in Sample 1 were administered a second CAPS, 2 to 3 days after the first one, by a third clinician. Sample 2 consisted of 24 veterans recruited for a research project on information processing in PTSD. All participants in Sample 2 were administered the CAPS twice, 2 to 3 days apart, by independent clinicians. For both Sample 1 and Sample 2, all raters were unaware of all other diagnostic information. For the dual administrations of the CAPS in Samples 1 and 2, a balanced incomplete blocks design with three raters was used. Two of the three raters independently interviewed each participant. All rater pairs interviewed the same number of participants, and rater order was counterbalanced.

Sample 3 consisted of 53 veterans and Sample 4 consisted of 67 veterans, all of whom were recruited for research projects on various aspects of the assessment of trauma and PTSD. Sample 5 consisted of 571 veterans seen for clinical services at the National Center between 1990 and 1996. For some analyses, we created a combined research sample, comprising Samples 1, 3, and 4, with a total sample of 243. We chose not to include the 24 participants from Sample 2 in the combined sample because they were recruited through a case-control rather than a naturalistic sampling scheme. Across all five samples, participants were primarily Caucasian (74–85%), primarily veterans of the Army (38–58%) and Marines (17–30%), and had at least some college education (64–92%). Mean age ranged from approximately 44 to 51 years. This range was influenced by the fact that the data were collected over a period of 6 years.

Measures

All participants in Sample 1 were administered the *DSM–III–R* versions of the CAPS and SCID PTSD module. In addition, all participants in Sample 3 and 507 of 571 participants (89%) of Sample 5 were administered the *DSM–III–R* version of the CAPS. All other participants were administered the *DSM–IV* version of the CAPS. The rating-scale anchors for the two versions of the CAPS are identical, which allowed us to combine participants who were administered different versions. It also allowed us to create PTSD diagnoses based on *DSM–III–R* and *DSM–IV* criteria for all participants, regardless of which version they were administered.

In order to do so, we had to consider three main changes in the PTSD criteria for *DSM–IV*. First, physiological reactivity was moved from the hyperarousal symptom cluster (Criterion D) to the reexperiencing cluster (Criterion B). Second, the definition of a traumatic event (Criterion A) was elaborated into a two-part definition, with A.1 requiring that the event involve life threat, serious injury, or threat to physical integrity, and A.2 requiring that the person experience intense fear, helplessness, or horror. Third, Criterion F, requiring clinically significant distress or functional impairment, was added.

In the present study, only one of these changes, moving physiological reactivity from Criterion D to Criterion B, was relevant, and thus we determined *DSM–III–R* versus *DSM–IV* diagnoses only this basis. The other two differences were essentially moot in the combat veterans we evaluated. First, regarding Criterion A, all participants had documented war-zone exposure in the Vietnam theater, and most had extensive exposure, having completed at least one 12- or 13-month tour of duty. Further, all those diagnosed with PTSD, even by the most lenient scoring rule, and most of those classified as non-PTSD, reported at least one specific event that would unequivocally satisfy Criterion A in either *DSM–III–R* or *DSM–IV*. Second, all veterans diagnosed with PTSD, as well as many of those classified as non-PTSD, reported significant distress or impairment (often both) associated with their symptoms, and therefore met Criterion F.

In both versions of the CAPS, information about distress and impairment is obtained from the intensity ratings for individual symptoms. In addition, both versions contain separate items explicitly assessing social and occupational impairment, although only the *DSM–IV* version contains an item explicitly assessing subjective distress.

In addition to the CAPS, participants also completed a battery of self-report measures that varied according to the purpose of their evaluation. In a concurrent validity analysis described below, we compared participants who met diagnostic criteria according to different CAPS scoring rules on the following self-report measures of PTSD, depression, anxiety, and global distress.

Mississippi Scale for Combat-Related PTSD (Mississippi Scale). The Mississippi Scale (Keane, Caddell, & Taylor, 1988) is the most widely used self-report measure of combat-related PTSD. It consists of 35 items, rated on a 5-point scale, based on the *DSM–III–R* PTSD criteria and associated features. It has demonstrated excellent psychometric properties in a growing number of investigations (D. W. King, King, Fairbank, Schlenger, & Surface, 1993; L. A. King & King, 1994; Kulka et al., 1991; McFall, Smith, Mackay, & Tarver, 1990). Keane et al. (1988) found an alpha of .94 and a 1-week test–retest reliability of .97. Regarding diagnostic use, they found that a cutoff of 107 had a sensitivity of .93, a specificity of .89, and an efficiency of .90 for predicting a consensus diagnosis of PTSD.

PTSD Checklist. The PTSD Checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993) is a 17-item scale originally based on the *DSM–III–R* PTSD criteria and revised in 1994 to correspond to the *DSM–IV* criteria. Using a 5-point scale, respondents indicate how much they were bothered by each PTSD symptom in the past month. In a sample of combat veterans, Weathers et al. (1993) found an alpha of .97 and test–retest reliability of .96. They also found that a cutoff of 50 had a sensitivity of .82, a specificity of .84, and a kappa of .64 against a SCID-based PTSD diagnosis. Investigating the PCL in a sample of motor vehicle accident victims, Blanchard, Jones-Alexander, Buckley, and Forneris (1996) reported an alpha of .94 and a correlation with the CAPS total severity score of .93. They further found that a cutoff of 44 had a sensitivity of .94, a specificity of .86, and an efficiency of .90.

PK scale of the Minnesota Multiphase Personality Inventory–2. The *PK* scale (Keane, Malloy, & Fairbank, 1984) has also been used widely in the assessment of combat-related PTSD. The original *PK* scale was empirically derived from the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1983) and it consisted of 49 MMPI items that best discriminated Vietnam combat veterans with and without PTSD. When the MMPI-2 (MMPI Restandardization Committee, 1989) was developed, three repeated items on the *PK* scale were dropped, reducing the number of items to 46, and one item was slightly reworded (see Lyons & Keane, 1992). In the MMPI-2 normative sample, alphas for the *PK* scale were .85 for men and .87 for women, and test–retest reliabilities were .86 for men and .89 for women (Graham, 1993). Keane et al. (1984) reported that a cutoff of 30 on the original 49-item version had an efficiency of .82 in two separate samples of Vietnam veterans. The diagnostic utility of the *PK* scale for assessing combat veterans has varied across subsequent investigations, due at least in part to variability in samples and diagnostic procedures, but in general has been supported. The PK scale has also been used successfully to assess civilian PTSD. Using a cutoff of 19, Koretzky and Peck (1990) found efficiencies of .87 and .88 in two samples of civilian trauma victims.

Beck Depression Inventory. The Beck Depression Inventory (BDI; Beck & Steer, 1993) is the most widely used self-report measure of depression. It consists of 21 items, each containing four statements that reflect increasing severity of a given symptom of depression. The psychometric properties of the BDI have been examined extensively in clinical and nonclinical populations and have been the subject of several review articles (e.g., Beck, Steer, & Garbin, 1988). The accumulated evidence strongly supports the BDI as a reliable and valid measure of the severity of current depression.

Beck Anxiety Inventory. The Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988) is a 21-item self-report measure of anxiety. Items consist of brief statements describing symptoms of anxiety, and they are rated on a 4-point scale. Beck and Steer (1993) reported alphas consistently above .90 across different samples and a 1-week test–retest reliability of .75. They also reported extensive evidence supporting the validity of the BAI as a measure of the severity of current anxiety.

Global Severity Index of the Symptom Checklist 90—Revised. The Symptom Checklist 90—Revised (SCL-90–R; Derogatis, 1992) is a 90-item self-report measure of psychopathology that assesses nine symptom dimensions (somatization, obsessive–compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism). Items consist of brief descriptions of symptoms and are rated on a 5-point scale. The SCL-90–R also yields three global scores, including the Global Severity Index (GSI), which is the mean severity score over all 90 items. As such, the GSI is a measure of overall psychological distress and is recommended for situations when a single summary score for the SCL-90–R is desired (Derogatis, 1992).

CAPS Scoring Rules

We examined the psychometric properties of nine scoring rules for converting CAPS frequency and intensity scores into a dichotomous PTSD diagnosis. The first four rules were rationally derived and the last five were empirically derived. For five of the scoring rules (Frequency ≥ 1/Intensity ≥ 2; Item Severity ≥ 4; Total Severity ≥ 45; Total Severity ≥ 65; Frequency ≥ 1/Intensity ≥ 2/Total Severity ≥ 65), a PTSD diagnosis can be constructed from the brief descriptions provided below. For four of the rules (Clinician-Rated 60, Clinician-Rated 75, SCID Diagnosis-Calibrated, and SCID Symptom-Calibrated), the CAPS item cutoffs required to generate a PTSD diagnosis are presented in the Appendix. For all scoring rules that involve dichotomizing individual CAPS items, a PTSD diagnosis is derived by first dichotomizing the items, and then following the *DSM–III–R* or *DSM–IV* algorithm for PTSD (one reexperiencing symptom, three avoidance and numbing symptoms, and two hyperarousal symptoms).

Frequency > 1/Intensity ≥ 2 (F1/I2). This was the original scoring rule proposed by Blake et al. (1990). According to this rule, a PTSD symptom is considered present if the frequency of the corresponding CAPS item is rated as 1 or higher and the intensity is rated as a 2 or higher. This roughly corresponds to Blanchard et al.'s (1995) more inclusive Rule of 3, the difference being that Blanchard et al. also considered a symptom to be present when the frequency was 2 or higher and the intensity was 1 or higher. That is, they considered a symptom to be present when the severity of the corresponding CAPS item (frequency + intensity) was 3 or higher.

Item Severity ≥ 4 (ISEV4). According to this rule, a PTSD symptom is considered present if the severity of the corresponding CAPS item is 4 or higher. This is the same as Blanchard et al.'s (1995) Rule of 4.

Clinician-Rated 60 (CR60). To develop this rule, a group of 25 clinicians with extensive PTSD experience rated every combination of frequency and intensity ratings for every item on the CAPS as *absent, subthreshold,* or *present.* According to this rule, a PTSD symptom is considered present if the combination of frequency and intensity for the corresponding CAPS item was rated as *present* by at least 60% of the clinicians.

Clinician-Rated 75 (CR75). This rule is based on the same ratings as the CR60 rule, except that a PTSD symptom is considered present if the combination of frequency and intensity for the corresponding CAPS item was rated as *present* by at least 75% of the clinicians.

SCID Diagnosis-Calibrated (DXCAL). This is an empirically derived rule based on data from Sample 1. Using Kraemer's (1992) methodology, we identified for each CAPS item the optimally efficient severity score (frequency + intensity) for predicting a SCID-based PTSD diagnosis. We

then used these optimally efficient severity scores as cutoffs for dichotomizing CAPS items. According to this rule, a PTSD symptom is considered present if the severity score for the corresponding CAPS item is greater than or equal to the empirically derived cutoff for that item.

SCID Symptom-Calibrated (SXCAL). This rule is similar to the DXCAL rule, except that for each CAPS item we identified the optimally efficient severity score for predicting the presence or absence of the corresponding SCID PTSD symptom. Thus, what distinguishes these two rules is that for the DXCAL we used the SCID-based PTSD diagnosis as the criterion for determining the optimal CAPS item cutoffs, whereas for the SXCAL we used the corresponding SCID PTSD item as the criterion.

Total Severity ≥ 45 (TSEV45). This is an empirically derived rule based on the total CAPS severity score (frequency + intensity summed across all 17 PTSD symptoms). Orr (1997) identified a total CAPS severity score of 45 as having the greatest concordance with physiological reactivity to script-driven imagery in adult female survivors of childhood sexual abuse.

Total Severity ≥ 65 (TSEV65). This is similar to the TSEV45 rule. Weathers et al. (1998) found a total severity score of 65 or higher to be the optimally efficient cutoff for predicting a PTSD diagnosis based on the SCID.

Frequency ≥ 1/Intensity ≥ 2/Total Severity ≥ 65 (F1/I2/TSEV65). This rule combines the F1/I2 and TSEV65 rules. It is intended to ensure both a significant overall level of PTSD symptom severity and a distribution of symptoms corresponding to *DSM–IV* diagnostic criteria.

Table 2
Kappa Coefficients Indicating the Reliability (Stability and Rater Equivalence) of Posttraumatic Stress Disorder (PTSD) Diagnoses Derived From Nine Clinician Administered PTSD Scale (CAPS) Scoring Rules

Scoring rule	Sample 1 (n = 60)	Sample 2 (n = 24)
Rationally derived rules		
Frequency ≥ 1/Intensity ≥ 2[a]	.81	.68
Item Severity ≥ 4	.82	.88
Clinician-Rated 60	.80	1.00
Clinician-Rated 75	.76	.83
Empirically derived rules		
SCID Diagnosis-Calibrated	.72	.78
SCID Symptom-Calibrated[a]	.89	1.00
Total severity ≥ 45	.85	.78
Total severity ≥ 65[a]	.86	1.00
Frequency ≥ 1/Intensity ≥ 2/ Total severity ≥ 65	.90	1.00

Note. Kappas are based on two administrations of the CAPS by independent raters. SCID = Structured Clinical Interview for *DSM–III–R*.
[a] Data in row were presented in Weathers et al. (1998).

Results

For our initial analysis we calculated kappa coefficients comparing PTSD diagnoses based on *DSM–III–R* versus *DSM–IV* criteria. Kappas for all nine scoring rules were at or very near unity in both the combined research sample (.97–1.00) and the clinical sample (.95–1.00), indicating a perfect or nearly perfect correspondence between *DSM–III–R* and *DSM–IV* criteria. Because the two versions of the *DSM* yielded essentially identical results, we used only *DSM–IV* criteria for all other analyses.

Table 2 presents kappa coefficients indicating the reliability of the different scoring rules based on two independent administrations of the CAPS in Samples 1 and 2. Because the design of the reliability study involved different occasions and different raters (i.e., test–retest with alternate forms), these kappas are more precisely referred to as coefficients of stability and rater equivalence (see Crocker & Algina, 1986). In Sample 1, the range of kappas was .72 for the DXCAL rule to .90 for the F1/I2/TSEV65 rule, indicating good to excellent reliability. In Sample 2, the kappas were somewhat more variable, ranging from .68 for the F1/I2 rule to 1.00 for the CR60, SXCAL, TSEV65, and F1/I2/TSEV65 rules. The kappas in Sample 2 corroborate those in Sample 1, and in several cases indicate stronger, even perfect, reliability. However, the Sample 1 kappas likely provide more stable estimates of reliability, in that the Sample 2 kappas may have been influenced by the case-control sampling scheme and the relatively small sample size. Kappa coefficients for individual CAPS items for the scoring rules involving individual items are available on request from Frank W. Weathers.

Table 3 presents data on the diagnostic utility of the nine scoring rules for predicting a PTSD diagnosis based on the SCID. These data are from Sample 1, in which all participants were administered the SCID PTSD module as well as at least one CAPS. The key comparisons among the rules pertain to the three kappa coefficients shown in Table 3. According to Kraemer (1992), the main reason for focusing on these kappa coefficients, which she refers to as quality indices, is that commonly reported measures of diagnostic utility, such as sensitivity, specificity, efficiency, and positive and negative predictive value, are uncalibrated measures of test performance that do not take into account chance agreement between test and diagnosis. The three quality indices, on the other hand, are calibrated such that a kappa of .00 indicates chance agreement between the test and the diagnosis, and a kappa of 1.00 indicates perfect agreement.

According to Kraemer (1992), $\kappa(1)$, representing the quality of sensitivity, ranges from .00, when sensitivity equals the level of the test (i.e., the proportion of test positives), to 1.00 when sensitivity is perfect. Representing the quality of specificity, $\kappa(0)$, ranges from .00, when specificity equals the complement of the level of the test (i.e., 1 − level of the test), to 1.00, when specificity is perfect. The third quality index, $\kappa(.5)$, which is the same as Cohen's kappa, represents the quality of efficiency. It is the most familiar of the three kappas, and typically is the only index of test quality presented in diagnostic utility analyses. A weighted average of $\kappa(1)$ and $\kappa(0)$, $\kappa(.5)$ ranges from .00, when efficiency equals chance agreement between test and diagnosis, to 1.00 when efficiency is perfect. Kraemer (1992) further demonstrated that the quality of positive predictive value equals the quality of specificity, and the quality of negative predictive value equals the quality of sensitivity.

As shown in Table 3, the highest values of $\kappa(.5)$ were obtained for the SXCAL, DXCAL, and F1/I2/TSEV65 rules, indicating that these were the optimally efficient rules and therefore the most valuable for differential diagnosis. The highest values of $\kappa(1)$ were obtained for the TSEV45, SXCAL, and DXCAL rules, indicating that these were the optimally sensitive rules and therefore most valuable for screening. The highest values of $\kappa(0)$ were obtained for the CR75, F1/I2/TSEV65, and CR60 rules, indicating that these were the optimally specific rules and therefore most valuable for confirming a diagnosis.

Table 3
Diagnostic Utility of Nine Clinician-Administered Posttraumatic Stress Disorder (PTSD) Scale Scoring Rules Versus a Structured Clinical Interview for DSM–III–R (SCID)-Based PTSD Diagnosis (N = 123, Base Rate = 54%)

Scoring rule	Level of test	Sensitivity	Specificity	PPV	NPV	Efficiency	κ(0)	κ(.5)	κ(1)
Rationally derived rules									
Frequency ≥ 1/Intensity ≥ 2[a]	.63	.91	.71	.79	.87	.82	.54	.63	.76
Item Severity ≥ 4	.61	.90	.73	.80	.85	.82	.56	.64	.73
Clinician-Rated 60	.43	.73	.93	.92	.74	.82	.83	.65	.53
Clinician-Rated 75	.39	.70	.98	.98	.73	.83	.95	.67	.51
Empirically derived rules									
SCID Diagnosis-Calibrated	.58	.91	.82	.86	.88	.87	.69	.74	.79
SCID Symptom-Calibrated[a]	.57	.91	.84	.87	.89	.88	.72	.75	.79
Total severity ≥ 45	.63	.93	.71	.79	.89	.83	.55	.65	.80
Total severity ≥ 65[a]	.49	.82	.91	.92	.81	.86	.82	.72	.65
Frequency ≥ 1/Intensity ≥ 2/ Total severity ≥ 65	.48	.82	.93	.93	.81	.87	.85	.74	.66

Note. Data are from Sample 1. Level of test = proportion of test positives; PPV = positive predictive value; NPV = negative predictive value; κ(0) = kappa coefficient representing quality of specificity; κ(.5) = kappa coefficient representing quality of efficiency; κ(1) = kappa coefficient representing quality of sensitivity.
[a] Data in row were presented in Weathers et al. (1998).

Table 4 presents the prevalence estimates of PTSD based on the nine scoring rules. As expected, the rules yielded a wide range of prevalence estimates in both the research (26–49%) and clinical (47–82%) samples. Although the rank order of the rules varied somewhat across the research and clinical samples, the F1/I2, ISEV4, and TSEV45 rules were the most lenient (yielding the highest prevalence estimates), and the F1/I2/TSEV65, CR60, and CR75 were the most stringent (yielding the lowest prevalence estimates). The DXCAL, SXCAL, and TSEV65 rules were intermediate to the others.

Finally, following Blanchard et al. (1995), we examined the impact of adopting increasingly stringent CAPS scoring rules. We created three groups of participants: (a) those who met diagnostic criteria for PTSD according to the CR75 rule, the most stringent rule we evaluated; (b) those who met criteria according to the TSEV65 rule, a moderate rule, but did not meet criteria according to the CR75 rule; and (c) those who met criteria according to the F1/I2 rule, a lenient rule, but did not meet criteria according to the two more stringent rules. As shown in Table 5, we compared these three groups on the Mississippi Scale, the PCL, the *PK* scale, the BDI, the BAI, and the GSI of the SCL-90–R. The PCL and the BAI were not included for the clinical sample as there were too few veterans who completed these measures as part of their clinical assessment. Also, the number of participants with complete data varied by instrument, as noted in Table 5.

Table 4
Prevalence Estimates of Posttraumatic Stress Disorder (PTSD) in Research and Clinical Samples as a Function of Clinician-Administered PTSD Scale Scoring Rule

Scoring rule	Sample: Combined research[a] (n = 243)	Sample: Clinical[b] (n = 571)
Rationally derived rules		
Frequency ≥ 1/Intensity ≥ 2	47.7	81.6
Item severity ≥ 4	45.3	78.1
Clinician-Rated 60	31.3	58.5
Clinician-Rated 75	25.9	47.3
Empirically derived rules		
SCID Diagnosis-Calibrated	43.2	73.4
SCID Symptom-Calibrated	41.6	69.7
Total severity ≥ 45	48.6	76.9
Total severity ≥ 65	34.2	59.7
Frequency ≥ 1/Intensity ≥ 2/ Total severity ≥ 65	33.7	58.7

Note. Values represent the percentage of the sample assigned a diagnosis of PTSD under each scoring rule.
[a] Comprises Samples 1, 3, and 4. [b] Sample 5.

Although this analysis included measures of anxiety, depression, and global distress, it was not intended as an investigation of the convergent and discriminant validity of the CAPS, an issue we have examined thoroughly elsewhere (see Weathers et al., 1998). Rather, like Blanchard et al. (1995), we simply sought to demonstrate that increasingly stringent CAPS scoring rules identify individuals with more severe PTSD and associated distress and impairment. It appears that the various CAPS scoring rules, ordered from most lenient to most stringent, reflect a dimension of PTSD severity, such that subgroups identified by different rules vary quantitatively rather than qualitatively with respect to their level of psychopathology.

As shown in Table 5, the three subgroups were rank ordered in the expected pattern on all of the measures in both the research and clinical samples. The CR75 group had significantly higher scores on all measures relative to the F1/I2 group. The TSEV65 group was intermediate to the other two groups, with significantly higher scores relative to the F1/I2 group in all but one instance, and lower, and sometimes significantly lower, scores relative to the CR75 group.

Although the pattern of results was as predicted, the effect sizes for some of the measures were modest. This was particularly the case for the clinical sample, most likely due to the restricted range of scores in these treatment-seeking veterans. Interestingly, the largest effect sizes were for the Mississippi Scale in the clinical

Table 5
Concurrent Validity of Three Clinician-Administered Posttraumatic Stress Disorder Scale Scoring Rules

Sample and scale	Scoring rule			
	F1/I2	TSEV65	CR75	eta^2
Clinical				
Mississippi Scale	110.53 (87)[a]	123.20 (90)[b]	129.44 (228)[c]	.159
MMPI-2 *PK*	80.99 (70)[a]	88.95 (81)[b]	92.06 (209)[b]	.085
BDI	22.31 (81)[a]	26.85 (86)[b]	30.95 (211)[c]	.096
SCL-90–R GSI	1.58 (74)[a]	1.78 (78)[a]	2.15 (207)[b]	.086
Combined research				
Mississippi Scale	98.74 (27)[a]	110.31 (16)[b]	114.33 (51)[b]	.235
PCL	47.91 (33)[a]	60.20 (20)[b]	67.98 (62)[c]	.407
MMPI-2 *PK*	71.52 (33)[a]	83.10 (21)[b]	90.13 (63)[b]	.238
BDI	18.97 (32)[a]	26.71 (21)[b]	30.08 (60)[b]	.182
BAI	14.64 (33)[a]	24.20 (20)[b]	29.00 (59)[b]	.235
SCL-90–R GSI	1.00 (24)[a]	1.83 (15)[b]	2.15 (56)[b]	.335

Note. Values represent means, with number of available cases in parentheses. Values whose superscripts differ are significantly different from one another at the .05 level. F1/I2 = Frequency ≥ 1/Intensity ≥ 2; TSEV65 = Total Severity ≥ 65; CR75 = Clinician-Rated 75; BDI = Beck Depression Inventory; MMPI-2 *PK* = Minnesota Multiphasic Personality Inventory–2 PK Scale *T* score; MMPI-2 *ANX* = MMPI-2 *Anxiety* Content Scale *T* score; SCL-90–R GSI = SCL-90–R Global Severity Index raw score; PCL = Posttraumatic Stress Disorder Checklist; BAI = Beck Anxiety Inventory.

sample and the PCL in the research sample. This could be seen as evidence of convergent validity, suggesting that there may be some specificity of the relationship between increasingly stringent scoring rules on the CAPS and severity of PTSD, as opposed to severity of depression, anxiety, or global distress. On the other hand, in the research sample the effect sizes for the BAI and GSI met or exceeded that of the Mississippi Scale. Further, the strong effect size found for the PCL could be due in part to the fact that the PCL, like the CAPS, contains items that precisely correspond to the *DSM–IV* criteria for PTSD.

Discussion

In this article, we described nine scoring rules for converting CAPS frequency and intensity scores into dichotomous PTSD diagnoses and compared these rules in terms of their reliability, diagnostic utility, and estimated prevalence of PTSD. We also examined the impact of adopting increasingly stringent rules on other indicators of PTSD and psychopathology. Finally, we examined the impact of using *DSM–III–R* versus *DSM–IV* diagnostic criteria for PTSD.

All nine rules demonstrated good to excellent reliability across two independent administrations of the CAPS. Diagnostic utility analyses revealed some variability among the rules in their quality of efficiency, although most were in the adequate to very good range. Greater variability among the rules was observed in their quality of sensitivity and specificity, indicating that some rules are more suitable for screening, and others are more suitable for confirming a diagnosis. As expected, we found that the nine rules yielded a wide range of prevalence estimates across both research and clinical samples, and thus could be characterized as ranging from relatively lenient (yielding high prevalence estimates) to relatively stringent (yielding low prevalence estimates). We also found that the choice of a CAPS scoring rule had important implications for the clinical status of those identified as PTSD positive: Participants who met diagnostic criteria for PTSD according to a stringent scoring rule had significantly higher scores on self-report measures of PTSD, depression, anxiety, and global distress relative to those who met criteria according a lenient rule.

These findings mirror those of Blanchard et al. (1995), who obtained PTSD prevalence estimates ranging from 27% to 44% for three CAPS scoring rules in a sample of motor vehicle accident victims. Blanchard et al. also found greater subjective distress and functional impairment in participants who met PTSD according to the most stringent scoring rule. Although the rules they evaluated differ somewhat from those used in the present study, both studies illustrate the substantial impact that using different CAPS scoring rules has on PTSD prevalence and severity of psychopathology in those identified as PTSD positive.

Finally, we found that the *DSM–III–R* and *DSM–IV* diagnostic criteria for PTSD yielded nearly identical results. This is not surprising, given that the *DSM–IV* revisions of the PTSD criteria were relatively minor, but this is one of the first studies to examine this issue empirically. A practical implication of this finding is that PTSD assessments conducted with the original version of the CAPS (based on *DSM–III–R* criteria) could be rescored according to *DSM–IV* criteria, with negligible impact on diagnostic status among those assessed.

These findings highlight the potential complexity and ambiguity involved in developing, evaluating, and selecting scoring rules for converting continuous severity scores into a dichotomous diagnosis. Any dimensional interview can be scored a number of different ways, and different scoring rules can yield markedly different outcomes. Dimensional interviews provide more options, but add a layer of complexity to the assessment process. We believe it is incumbent on test developers to propose and empirically evaluate different scoring rules for dimensional instruments and to develop empirically based recommendations for test users. In turn, it is incumbent on test users to select the most appropriate scoring rule

for a given assessment task and to explicitly identify and defend their choice. For example, it is insufficient for an investigator to report only that PTSD diagnoses were made on the basis of the CAPS, although such limited descriptions are common in the literature. A complete operational definition would include the qualifications and training of the interviewers, the circumstances under which the interview was administered, the version of the CAPS that was used, the scoring rule that was applied to obtain a diagnosis, and a justification linking the choice of scoring rule to the purpose of the assessment.

Regarding the best scoring rules for the CAPS, it is premature to make firm recommendations without cross-validation in other trauma populations and settings. At this point, whenever feasible, the best strategy may be to use several different scoring rules and evaluate the impact of the various rules on the outcome of a study. However, when such a strategy is not feasible, some general guidelines may be followed. For screening (i.e., when false negatives are to be avoided), a lenient rule such as the F1/I2 rule would be appropriate. For confirming a diagnosis or creating a homogeneous group of individuals with unequivocal PTSD (i.e., when false positives are to be avoided), a stringent rule such as F1/I2/SEV65 or CR60 would be appropriate. For differential diagnosis, when false positives and false negatives are equally undesirable, a moderate rule such as SXCAL would be a reasonable choice.

One limitation of this study is that it includes only male Vietnam theater veterans, most of whom were seeking some type of services from the Boston Veterans Affairs Medical Center. A second limitation is that the diagnostic utility analyses were conducted using a SCID-based PTSD diagnosis as the gold standard. According to Kraemer (1992), in the evaluation of the quality of a test, the performance of the test is limited by the reliability of the gold standard. Thus, a good test may appear to be of poor quality simply because the gold standard is unreliable. She argues that the kappa indicating the reliability of the gold standard is an essential benchmark for evaluating the quality of a test. Tests with quality indexes that approach or exceed the kappa for the current gold standard may be good candidates to supplant it as the new criterion.

The SCID PTSD module has been used as a criterion measure in psychometric studies of other PTSD instruments, but it has not been subjected to a rigorous psychometric evaluation itself. There is some evidence to suggest that the SCID PTSD module may be less reliable than the CAPS and some of the other dimensional PTSD interviews. For example, Keane et al. (1998) found a kappa of .68 when the SCID PTSD module was administered twice by independent clinicians. This value is substantially lower than the most reliable CAPS rules reported in this study, and lower than even the least reliable CAPS rules. Further, this value is lower than the kappa indicating the quality of efficiency for four of the nine scoring rules evaluated in this study. In sum, the CAPS may be more reliable than the SCID PTSD module and may be more predictive of the SCID than the SCID is of itself. Future studies could test these hypotheses directly by evaluating the reliability of the SCID PTSD module, the reliability of the CAPS, and the diagnostic use of the CAPS against the SCID in the same sample.

In conclusion, this article illustrates the impact of adopting different scoring rules for the CAPS and the importance of specifying and justifying a particular rule for a given PTSD assessment task. More studies are needed to determine the generalizability of our findings across other trauma populations and other settings. The issues and methods we have described are broadly applicable to any structured interview, for PTSD or any other disorder, that uses dimensional rather than categorical rating scales to evaluate symptom severity.

References

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

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.

Beck, A. T., & Steer, R. A. (1993). *Beck Depression Inventory manual.* San Antonio, TX: The Psychological Corporation.

Beck, A. T., Steer, R. A., & Garbin, M. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100.

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.

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., Hickling, E. J., Taylor, A. E., Forneris, C. A., Loos, W., & Jaccard, J. (1995). Effects of varying scoring rules of the Clinician-Administered PTSD Scale (CAPS) for the diagnosis of post-traumatic stress disorder in motor vehicle accident victims. *Behavior Research and Therapy, 33,* 471–475.

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,* 669–673.

Blashfield, R. K. (1984). *The classification of psychopathology: Neo-Kraepelinian and quantitative approaches.* New York: Plenum.

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

Davidson, J. R. T., Smith, R. D., & Kudler, H. S. (1989). Validity and reliability of the *DSM–III* criteria for posttraumatic stress disorder: Experience with a structured interview. *Journal of Nervous and Mental Disease, 177,* 336–341.

Derogatis, L. R. (1992). *SCL-90-R: Administration, scoring & procedures manual—II (2nd ed.).* Towson, MD: Clinical Psychometric Research.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM–IV Axis I Disorders—Clinician Version (SCID–CV).* Washington, DC: American Psychiatric Press.

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.

Graham, J. R. (1993). *MMPI-2: Assessing personality and psychopathology (2nd ed.).* New York: Oxford University Press.

Hathaway, S. R., & McKinley, J. C. (1983). *The Minnesota Multiphasic Personality Inventory manual.* New York: Psychological Corporation.

Herman, D. S., Weathers, F. W., Litz, B. T., & Keane, T. M. (1996). The PK scale of the MMPI-2: Reliability and validity of the embedded and stand-alone versions. *Assessment, 3,* 437–442.

Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi Scale for combat-related posttraumatic stress disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology, 56,* 85–90.

Keane, T. M., Kolb, L. C., Kaloupek, D. G., Orr, S. P., Blanchard, E. B., Thomas, R. G., Hsieh, F. Y., & Lavori, P. W. (1998). Utility of psychophysiological measurement in the diagnosis of posttraumatic stress disorder: Results from a Department of Veterans Affairs cooper-

ative study. *Journal of Consulting and Clinical Psychology, 66,* 914–923.

Keane, T. M., Malloy, P. F., & Fairbank, J. A. (1984). Empirical development of an MMPI subscale for the assessment of combat-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 52,* 888–891.

King, D. W., King, L. A., Fairbank, J. A., Schlenger, W. E., & Surface, C. R. (1993). Enhancing the precision of the Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: An application of item response theory. *Psychological Assessment, 5,* 457–471.

King, D. W., Leskin, G. A., King, L. A., & Weathers, F. W. (1998). Confirmatory factor analysis of the Clinician-Administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment, 10,* 90–96.

King, L. A., & King, D. W. (1994). Latent structure of the Mississippi Scale for Combat-Related Post-traumatic Stress Disorder: Exploratory and higher order confirmatory factor analyses. *Assessment, 1,* 275–291.

Koretzky, M. B., & Peck, A. H. (1990). Validation and cross-validation of the PTSD Subscale of the MMPI with civilian trauma victims. *Journal of Clinical Psychology, 46,* 296–300.

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

Kulka, R. A., Schlenger, W. E., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. R., & Weiss, D. S. (1991). Assessment of posttraumatic stress disorder in the community: Prospects and pitfalls from recent studies of Vietnam veterans. *Psychological Assessment, 3,* 547–560.

Litz, B. T., Schlenger, W. E., Weathers, F. W., Cadell, J. M., Fairbank, J. A., & LaVange, L. M. (1997). Predictors of emotional numbing in posttraumatic stress disorder. *Journal of Traumatic Stress, 10,* 607–618.

Lorr, M. (1986). Classifying psychotics: Dimensional and categorical approaches. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Toward the DSM–IV* (pp. 331–345). New York: Guilford Press.

Lyons, J. A., & Keane, T. M. (1992). Keane PTSD scale: MMPI and MMPI-2 update. *Journal of Traumatic Stress, 5,* 1992.

McFall, M. E., Smith, D. E., Mackay, P. W., & Tarver, D. J. (1990). Reliability and validity of Mississippi Scale for Combat-Related Posttraumatic Stress Disorder. *Psychological Assessment, 2,* 114–121.

MMPI Restandardization Committee. (1989). *MMPI-2: Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Orr, S. P. (1997). Psychophysiologic reactivity to trauma-related imagery in PTSD. In R. Yehuda & A. C. McFarlane (Eds.), *Annals of the New York Academy of Sciences, Vol. 821: Psychobiology of posttraumatic stress disorder* (pp. 114–124). New York: The New York Academy of Sciences.

Orsillo, S., Weathers, F. W., Litz, B. T., Steinberg, H. R., Huska, J. A., & Keane, T. M. (1996). Current and lifetime psychiatric disorders among veterans with war-zone-related post-traumatic stress disorder. *Journal of Nervous and Mental Disease, 184,* 307–313.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). *Structured Clinical Interview for DSM–III–R.* Washington, DC: American Psychiatric Press.

Weathers, F. W., Blake, D. D., Krinsley, K. E., Haddad, W., Ruscio, A. M., Keane, T. M., & Huska, J. A. (1998). The reliability and validity of the Clinician-Administered PTSD Scale (CAPS). Manuscript submitted for publication.

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 presented at the annual meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

Weathers, F. W., Litz, B. T., Herman, D. S., Keane, T. M., Steinberg, H. R., Huska, J. A., & Kraemer, H. C. (1996). The utility of the SCL-90-R for the diagnosis of war-zone-related PTSD. *Journal of Traumatic Stress, 9,* 111–128.

Widiger, T. A. (1997). Mental disorders as discrete clinical conditions; Dimensional versus categorical classification. In S. M. Turner & M. Hersen (Eds.), *Adult psychopathology and diagnosis* (3rd ed.; pp. 3–23). New York: Wiley.

Appendix

Item Cutoffs for Generating a Posttraumatic Stress Disorder Diagnosis According to Four Different Scoring Rules for the Clinician-Administered Posttraumatic Stress Disorder Scale

Table A1
Frequency–Intensity Pairs for Dichotomizing Clinician-Administered Posttraumatic Stress Disorder Scale (CAPS) Items According to the Clinician-Rated 60 Scoring Rule

CAPS item	Frequency–intensity pairs
1	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
2	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
3	1-3,1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
4	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
5	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
6	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
7	2-3,2-4,3-3,3-4,4-2,4-3,4-4
8	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
9	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
10	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
11	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
12	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
13	1-3,1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
14	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
15	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
16	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
17	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4

Note. Values represent the frequency–intensity combinations that indicate the presence of a symptom, according to the Clinician-Rated 60 scoring rule. For a given CAPS item, if an individual's frequency and intensity scores match one of the frequency–intensity pairs listed, that item is counted as a symptom toward a PTSD diagnosis.

Table A2
Frequency–Intensity Pairs for Dichotomizing Clinician-Administered Posttraumatic Stress Disorder Scale (CAPS) Items According to the Clinician-Rated 75 Scoring Rule

CAPS item	Frequency–intensity pairs
1	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
2	2-3,2-4,3-3,3-4,4-2,4-3,4-4
3	1-4,2-3,2-4,3-3,3-4,4-2,4-3,4-4
4	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
5	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
6	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
7	2-4,3-3,3-4,4-3, 4-4
8	2-4,3-3,3-4,4-2,4-3,4-4
9	2-4,3-3,3-4,4-2,4-3,4-4
10	2-4,3-3,3-4,4-2,4-3,4-4
11	2-4,3-3,3-4,4-2,4-3,4-4
12	2-4,3-2,3-3,3-4,4-2,4-3,4-4
13	2-3,2-4,3-3,3-4,4-2,4-3,4-4
14	2-4,3-3,3-4,4-2,4-3,4-4
15	2-4,3-2,3-3,3-4,4-2,4-3,4-4
16	1-4,2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4
17	2-3,2-4,3-2,3-3,3-4,4-2,4-3,4-4

Note. Values represent the frequency–intensity combinations that indicate the presence of a symptom, according to the Clinician-Rated 75 scoring rule. For a given CAPS item, if an individual's frequency and intensity scores match one of the frequency–intensity pairs listed, that item is counted as a symptom toward a PTSD diagnosis.

Table A3
Severity Score Cutoffs for Dichotomizing Clinician-Administered Posttraumatic Stress Disorder (PTSD) Scale (CAPS) Items According to the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders *(3rd ed., rev.; SCID) Diagnosis-Calibrated and SCID Symptom-Calibrated Scoring Rules*

	Scoring rule	
CAPS item	SCID diagnosis-calibrated	SCID symptom-calibrated
1	3	3
2	3	2
3	3	3
4	3	3
5	4	4
6	4	4
7	4	5
8	5	5
9	6	5
10	3	6
11	4	5
12	4	4
13	5	4
14	3	4
15	6	3
16	3	3
17	3	3

Note. Values represent severity score cutoffs that indicate the presence of a symptom, according to the SCID diagnosis-calibrated and SCID symptom-calibrated scoring rules. For a given CAPS item, if an individual's severity score is greater than or equal to the value listed, that item is counted as a symptom toward a PTSD diagnosis.

Received July 27, 1998
Revision received January 12, 1999
Accepted January 14, 1999 ■