INSTRUCTION MANUAL

National Center for PTSD

CLINICIAN-ADMINISTERED PTSD SCALE (CAPS)

Content

D. Blake, F. Weathers, L. Nagy, D. Kaloupek, G. Klauminzer, D. Charney, T. Keane, & T. C. Buckley.

National Center for Posttraumatic Stress Disorder

Behavioral Science Division - Boston Neurosciences Division - West Haven

November, 2000

CLINICIAN-ADMINISTERED PTSD SCALE (CAPS)

Description of the CAPS

The CAPS is a semi-structured interview that is designed to assess the essential features of *Posttraumatic Stress Disorder* as defined by the DSM-IV (American Psychiatric Association, 1994). It can also be used to assess associated features of the diagnostic syndrome (e.g., survivor guilt). In addition, the CAPS can also be used to assess the essential features of *Acute Stress Disorder* as currently defined by DSM-IV. The interview is designed to accommodate different time spans post-trauma as the referrent point for diagnosis. Specifically, the CAPS affords the clinician flexibility to inquire about symptoms and diagnostic status over the past week, most recent month, and/or for lifetime diagnosis. Any one, or all three, of the time frames may be used depending on the nature of the task at hand.

This most recent version of CAPS was designed with flexibility in mind. The nature of the instrument is such that it lends itself to a variety of clinical and research applications. For example, it may be used to provide categorical ratings of diagnostic status based on formal diagnostic criteria (i.e., PTSD present or absent). Alternatively, it can also provide a quantitative index of symptom severity for individual symptoms, symptom clusters, and/or the entire syndrome. Both frequency and intensity scores are derived for each individual symptom. As such, the CAPS can be used to track changes in diagnostic status and also to track changes in finer gradations of symptom severity over time, which is essential to both treatment outcome research and documenting status changes in clinical settings. The combination of both categorical and dimensional ratings also allows for flexibility in the application of data-analytic techniques when ratings of PTSD diagnosis and symptom status are the dependent variables in question.

In a series of psychometric studies the CAPS has been shown to a reliable psychometric instrument based on analyses of inter-rater reliability, test-retest reliability, and internal consistency. It is also the case that the CAPS is arguably the most valid measure of PTSD relative to other well validated structured interview and self-report instruments of PTSD (for detailed review of empirical work in this area, see the work of Weathers, Keane, & Davidson, 2000 in Appendix #1).

In addition to providing a means to evaluate the frequency and intensity of each symptom, the CAPS provides a means to evaluate: a) the impact of the symptoms on the patient's social and occupational functioning; b) the overall severity of the symptom complex; c) the patient's global improvement since baseline; and d) the validity of ratings obtained.

Because the CAPS was designed to meet multiple clinical and research goals, a variety of scoring algorithms are at the disposal of the assessing clinician depending on the nature of their task. For example, if the CAPS is being used for screening purposes, the assessor may wish to utilize a more “lenient” scoring algorithm when calculating the presence vs. absence of diagnosis based on

interview findings. Such a strategy minimizes false-negatives, which is very important for screening applications. Conversely, should the goal of the assessor be to minimize false positives, as might be the case in a treatment outcome study, he/she might then utilize a more stringent scoring rule to arrive at a diagnosis. The various scoring algorithms that have been studied with the CAPS are presented in Appendix-2 of this document.

The CAPS interview contains the following components:

Life Events Checklist: Criterion A Assessment
PTSD Symptoms
- Criterion B: Items 1-5
- Criterion C: Items 6-12
- Criterion D: Items 13-17

Lifetime Diagnosis Instructions
CAPS Global Ratings: Items 20-25
Associated or Hypothesized Features (optional): Items 26-30

Because CAPS users may have different purposes for its use, it is possible to customize both versions of the CAPS by eliminating less relevant components. However, this customization should be limited to changes which do not compromise the psychometric properties of the scale, such as adding or deleting scale anchor points.

General Instructions

The CAPS is a structured clinical interview designed to assess the seventeen symptoms for Post Traumatic Stress Disorder (PTSD) outlined in the DSM-IV, along with five associated features. It is to be used by mental health professionals who have experience conducting diagnostic interviews and who have formal knowledge of psychopathology and DSM-IV. This instruction manual should be studied and referred to by anyone who wishes to use the CAPS.

In general, all ratings should be based on the patient's report. However, the final rating should be made from the collective consideration of the patient's report, the interviewer's confidence in that report, and the patient's behavior during the interview process. The time frame for rating each item can be either the month prior to the interview , a one month period following the trauma (Lifetime Diagnosis) or the previous week (One Week Symptom Status Ratings).

Administration

Because the CAPS is designed to assess responses to specific traumatic events, it is essential for the clinician to inquire about possible Criterion A events that the patient may have been exposed to during their lifetime. Using the example narrative on Page 2 of the CAPS interview, the clinician should introduce the fact that they will be asking the patient about such events. The Life Events Checklist (LEC), should then be given to the patient so that he/she may mark those events that have occurred during their lifetime. After reviewing the LEC, the assessing clinician should inquire as to which events (up to three) were the “worst” in terms of emotional difficulties the patient experienced in their aftermath. Use the prompt questions on pages 2 and 3 to determine whether or not the essential features of the event met Criterion A as outlined in DSM-IV, in which case assessment for PTSD in relation to the event should take place. The CAPS interview can be conducted focusing on all events that meet Criterion A. For patients with multiple traumas, it may be the case that multiple interviews will be conducted, focusing exclusively on the response to one

trauma at a time.

A standard instructional set should be provided to the patient before starting the interview:

> **"We are interested in learning about the problems you may be experiencing. In order to do that, I am going to ask you questions about 22 different symptoms: For each symptom I want to find out if you've experienced it, and if so, about how often you've experienced it over the past ('week' or 'month' depending on time frame). Then, for each symptom that you've experienced, I want to find out how strong it was. I will also ask you about how these symptoms have affected your social life and work. Please try to keep your answers short and to the point. If I'm not sure I understand a problem you may be having, I will ask you more about it until I am sure.**
>
> **Do you have any questions?"**

Items should be rated based primarily on the patient's verbal response to the prompt questions or to comparable alternatives. If after asking the prompt questions, the patient's response is not sufficient to make a rating, the same question can be restated in terms which are more familiar to her or him. For example, **"Have you tried to avoid activities or situations that remind you of the event(s)?"** can be restated as **"Have you avoided doing things or being in places that reminded you of your experiences in Vietnam (or during the assault, the earthquake, etc.)?"** The interviewer can then ask additional questions to help select the most accurate descriptor from each rating continuum, e.g., **"Do you mean you experienced these symptoms every day or several times a week?"** Additional questions may involve asking about the reciprocal of the symptom in question; e.g., sleep disturbance can be gauged by asking the patient how many hours or how restful was his/her sleep each night during the time period in question. Clinical observation can also serve as a useful source of information, such as in assessing emotional numbing, social impairment, concentration, etc., as well as in determining the reliability of patient reporting.

The CAPS time frame can vary depending on the goals of the assessing clinician. . The time frame for the CAPS can be the one-month period preceding the interview (Current Diagnosis) or the most symptomatic one-month period following the traumatic event(s) (Lifetime Diagnosis). The time frame can also be be for the one-week period immediately preceding the interview (Symptom Status). Slight modifications in wording have to be made in order to accommodate this time frame differential, as can be seen in the time frame qualification occurring either before or after the prompt questions: Clinical judgement should be used in deciding the exact order or the phrasing used in the prompt questioning. For example, the CAPS interviewer may determine early in the interview that the patient has difficulty ascertaining symptom presence without first being given a time frame qualification. On the other hand the interviewer may find that the time frame qualification at the start of the prompt questions interferes with the patient's ability to decipher and respond to the questions. In either case, corresponding corrections are acceptable.

Ratings should be derived only from information obtained during the interview. The primary source from which ratings are made is the patient's responses to initial and follow-up questions. The interviewer must also use clinical judgment in making symptom ratings. For example, if the interviewer is concerned that a particular symptom is present, he or she should not allow a patient's denial of the symptom to go unchallenged. Follow-up questions should be used in establishing the most valid rating. If the patient provides unsolicited information at any time during the interview,

that additional information can also be used in making the ratings.

Space is provided for written descriptions of the interviewee's signs and symptoms. For each of the CAPS items there is a space below the "Frequency" prompt questions and response options for supplementing the ratings with a written description of critical parts of the patient's responses. This space contains the heading of "Description/Examples:" to designate an open section where the interviewer can note those aspects of the patient's response that were key in making the Frequency and Intensity ratings. For example, when asking about physiologic reactivity to events that symbolize aspects of the trauma (Item #5), an interviewer might write: "gets 'edgy' when in dark, wooded areas... ...heart pounds, very tense and keyed up" and "very upsetting, can't do much else but leave/wait to calm down." This section can also be used to record examples given by the patient, such as a description of what she or he considers to be a "flashback" experience (Item #3); this information can be helpful in determining whether the patient's symptom has the quality and intensity of this extreme dissociative experience. An exception to this open format for recording interview information is the structure provided for the item on sleep disturbance (Item #13): Here, due to the discrete quality of the problem, specific questions are asked in order to help in making a judgement about symptom intensity.

Responses that are thought to be inaccurate or invalid should be noted. If it is apparent that the patient's report is distorted or grossly inaccurate or otherwise not valid, the suspected invalidity of all involved items should be noted by circling the "QV" (Questionable Validity) underneath the the prompt questions; an aggregated rating of validity (item #23) should be made after the core PTSD symptoms have been assessed (see "Coding CAPS Ratings" below). When rating a response as "QV", the clinician should make a note in the space provided outlining his/her reason for doing so.

Ratings of Symptom Frequency

Frequency ratings are to be made on a 5-point continuum, from the lowest frequency (never or none of the time) to the highest (daily or all of the time). The interviewer should determine the most accurate rating along this continuum by first stating the prompt questions, and if necessary, comparable alternatives (i.e., stated in more colloquial terms), specific to each item.

	Frequency
Prompt questions	**Have you ever tried to stay away from activities or situations that reminded you of the event(s)? How often in the past month?**
Rating option continuum	**0 Never** **1 Once or twice** **2 Once or twice a week** **3 Several times a week** **4 Daily or almost every day**

If the prompt questions and their follow-ups do not produce the specificity required, the next step is to give the patient **response options** by naming those anchor point descriptors which appear to most closely reflect the patient's status (as indicated by his/her earlier responses to the prompt questions). On the other hand, all prompt questions need not be asked if an accurate rating is obtained with the initial prompt question.

Ratings of Symptom Intensity

Intensity ratings, which tap both **symptom intensity** and **degree of impairment**, are also to be

made on a 5-point scale, from the lowest intensity (none or no problem with symptom) to the highest (extreme, incapacitating). Again, the interviewer should first state the prompt questions and appropriate follow-up questions.

Intensity

Prompt question	**How much effort did you make to avoid activities or situations related to the event(s)? [rate all attempts at behavioral avoidance, e.g., combat veteran whoavoids veteran activities, war movies, etc.]**
Rating option continuum	**0 No effort** **1 Mild, minimal effort** **2 Moderate, some effort, avoidance definitely present** **3 Severe, considerable effort, marked avoidance** **4 Extreme, drastic attempts at avoidance**

If the prompt questions do not lead to a single, fitting rating, they can be followed by asking the patient to choose the most accurate rating/description among two or more possible options. If a symptom frequency is "0" (zero), the intensity rating by default is also coded a "0" and the interviewer should proceed to the next symptom.

Onset and Duration of Symptoms

Both questions #18 and #19 are fairly straightforward. As with other CAPS questions, the clinician can rephrase the questions should the patient not fully understand what is being asked of him or her. However, most patients will understand the phrasing as written on both of these questions.

CAPS Global Ratings

After rating each of the seventeen core PTSD symptoms, the interviewer should complete five questions which pertain to the patient's overall degree of impairment, his or her improvement since an earlier measurement, and the estimated validity of the patient's responses during the interview. As with the other CAPS ratings, the interviewer ratings are made on a 5-point scale. The items are as follows:

21. **Impact on Social Functioning:** Use prompt questions to determine the impact of PTSD symptoms on social functioning. Rated from 0 (no adverse impact) to 4 (extreme impact). For inpatients, ratings should be based on the patient's report of social activities and interactions in the hospital.

22. **Impact on Occupational Functioning:** Use prompt questions to determine the impact of PTSD symptoms on work and occupational functioning. Rated from 0 (no adverse impact) to 4 (extreme impact). For inpatients, ratings should be based on the interviewer's judgment of the extent to which the PTSD symptom complex has contributed to the patient's employment status during the time period in question.

23. **Global Validity:** Based in part on the number of QV's (questionable validity of item rating) circled on the interview form, the interviewer should estimate the validity of the set of symptom ratings obtained. The validity rating is recorded on a scale from 0 (excellent) to 4

(invalid responses).

24. **Global Severity:** Interviewer's judgment of the overall severity of the patient's illness. Rated from 0 (asymptomatic) to 4 (extreme symptoms, pervasive impairment).

25. **Global Improvement:** Rate total overall improvement present **since an earlier rating** whether or not it is due to a given treatment. Rated from 0 (asymptomatic) to 4 (no improvement).

Ratings of Associated or Hypothesized Features

Five additional symptoms are included in the CAPS to examine features associated with PTSD and those associated with Acute Stress Disorder (as is found in the DSM-IV). These items are optional, but can be useful for assessing changes associated with pharmacological and psychosocial treatments, and may further our general understanding of PTSD as a clinical entity. They should be rated in the same manner as the first 17 items (see Administration above).

Coding CAPS Ratings

Underneath each CAPS item is a qualifier with the initials "QV," which is an acronym for "Questionable Validity." To the right of the symptom questions are the initials “F” and “I”, which stand for "Frequency," and "Intensity." After obtaining ratings for a given symptom (i.e., circling the number to the left of the selected descriptor),the interviewer should write these numbers in the blank next to the "F" and "I". In cases where the interviewer has significant concern about the accuracy or veracity of the patient's response, "QV" should be circled and brief note as to why this was circled; this information will be used later in estimating the overall validity of the patient's responding during the CAPS interview.

For each symptom in the Current, Lifetime Diagnosis, and Weekly Status Version of the CAPS there is a double-column coding block to the right of the prompt questions. Under each block is a Yes/No rating which is meant for checking, if the minimum criteria for that symptom have been met (i.e., a "1" or greater for frequency and a "2" or greater for intensity). This practice will expedite the summarization which takes place after the interview.

Lifetime Status Prompts

Lifetime PTSD status may be assessed for those patients who do not currently meet the full set of diagnostic criteria. After assessing current symptoms, introduce the lifetime prompts with the following instructional set:

Lifetime Symptom Query

Has there been any time period since the trauma in which you were significantly more troubled than in the past month by the symptoms that I've just asked you about? NO YES

Did this period or these periods last for at least one month? **NO YES**

Approximately when did this/these period(s) begin and end? **____ to ____**
____ to ____
____ to ____

(For multiple time periods):

During which of these time periods were you most troubled by or experienced the greatest number of symptoms? **____ to ____**

After the above questions have been answered, the interviewer should go back to each of the CAPS items and inquire about the presence of each respective symptom during the time frame identified. The interviewer can proceed by using the following instructional set:

> **"During the month you identified as the worst time, did you experience (symptom)? How often did (symptom) occur?**

In this manner, the interviewer rates the symptom frequency rating for the time period in question, which is recorded in the appropriate space in the column to the right of the symptom ratings.

Finally, using the same method as that employed for the initial ratings, the interviewer determines the symptom intensity during the time period in question. This rating is also recorded in the designated space below the symptom rating (see above).

CAPS Summary Sheet

After the CAPS interview is complete, all of the resulting data should be coded onto the accompanying Summary Sheet. Summarizing the data ensures that the results are immediately usable. This practice also serves as a double-check for determining whether all CAPS items have been addressed and responses coded appropriately. Patient and interviewer identifying data can be copied directly from the CAPS cover sheet. The nature of the traumatic event(s) should then be summarized briefly to indicate whether Criterion A was met. Finally, all data coded in the boxes to the right of each item should be copied onto the summary sheet where indicated. Summary scores after each criterion should also be calculated at this point, the results of which are to be transcribed onto the appropriate places on the summary sheet.

Deriving Diagnostic Scores

Once the clinician has finished administering the CAPS and transferred the data to the Summary Sheet, he/she is then faced with the task of summarizing the data to arrive at a diagnostic decision. There are 9 different scoring rules for the CAPS that have been examined empirically. Depending on the application, different scoring rules may be called for (e.g., one that maximizes sensitivity might be best for screening applications). As such, the clinician should consult the 9 scoring rules

in Appendix-2 of this document to determine which rule is most appropriate to suit their needs.

Interviewer Preparation

1. Study the General Instructions, the CAPS Description, and the Ten CAPS Do's and Don't's in this manual.

2. Carefully read through every item of the CAPS, making sure that you understand all of the prompt questions.

3. Practice reading the CAPS prompt questions aloud until they sound natural to you.

4. Administer the CAPS with a colleague who can assume the role of a patient.

5. Practice using the CAPS on patients who are similar to those who will be assessed more formally. These can be joint interviews with each rater making independent ratings, followed by a discussion of the sources of disagreements in the ratings. If possible, audiotape the interview to facilitate the training process (e.g. by using tape playback to clarify patient responses).

Ten CAPS Do's and Don't's*

1. DO give the subject a brief explanation of the purpose of the interview before beginning. In research studies this will usually be part of obtaining informed consent.

   DON'T apologize for using a structured interview. ("I have to read these questions. Many of them won't apply to you. Just bear with me." "I have to give this standardized interview.")

2. DO make sure that you and the patient are focusing on the same (and the appropriate) time period for each question.

   DON'T assume that symptoms the patient is describing cluster together in time unless you have clarified the time period. For example, the patient may be talking about a symptom that occurred a year ago and another symptom that appeared last week, when you are focusing on concurrent symptoms that occurred during a one month period seven years ago.

3. DO stick to the initial prompt questions as they are written, except for necessary modifications to take into account what the patient has already said, his or her comprehension level, or to request elaboration or clarification.

   DON'T make up your own **initial** questions because you think you have a better way of getting at the same information. A lot of care has gone into the exact phrasing of each question.

4. DO make sure that the patient understands your question. It may be necessary to ask patients if they understand what you are asking about and to repeat or rephrase questions.

   DON'T use words that the patient does not understand.

5. DO feel free to ask additional clarifying questions, such as, "Do you experience sleep problems once a week or several times a week?"

   DON'T use the interview simply as a checklist, reading off all options from which the patient gives his or her endorsement.

6. DO use your judgment about a symptom, taking into account all of the information available to you, and confronting the patient (tactfully and diplomatically) about responses that are inconsistent with previously obtained information.

   DON'T necessarily accept a patient's response if it contradicts other information or you have reason to believe it is not valid.

* Based on corresponding section in Spitzer, R.L., Williams, J.B.W., Gibbon, M. & First, M.B. (1988). Instruction manual for the structured clinical interview for DSM-III-R (SCID, 6/1/88 Revision). New York: Biometrics Research Department of New York State Psychiatric Institute.

7. DO focus on obtaining the information necessary to judge all of the particulars of the criterion under consideration. This may require asking additional questions, and the interviewer should freely use the space allocated in the "Description/Examples:" sections.

   DON'T focus only on getting an answer to the CAPS prompts when additional questions are warranted.

8. DO proceed sequentially through the CAPS.

   DON'T skip over a section without filling anything in because you are certain that it does not apply.

9. DO select the rating options that are presented.

   DON'T make illegitimate ratings, such as a "3.5" when the choices are "3" or "4."

10. DO complete the CAPS Summary Sheet as soon as possible after the interview is completed.

DON'T delay determining the patient's symptom or diagnostic status when that information may be important.

Appendix #1

Psychometric Properties of the Clinician Administered Posttraumatic Stress Disorder Scale.

Adapted from: Weathers, F.W., Keane, T.M., & Davidson, J.R.T. (In Press). The Clinician Administered PTSD Scale: A Review of the First Ten Years of Research, *Depression and Anxiety,*

Abstract

The Clinician-Administered PTSD Scale (CAPS) is a structured interview for assessing posttraumatic stress disorder (PTSD) diagnostic status and symptom severity. In the 10 years since it was developed, the CAPS has become a standard criterion measure in the field of traumatic stress and has now been used in more than 200 studies. In this paper we first trace the history of the CAPS and provide an update on recent developments. Then we review the empirical literature, summarizing and evaluating the findings regarding the psychometric properties of the CAPS. The research evidence indicates that the CAPS has excellent reliability, yielding consistent scores across items, raters, and testing occasions. There is also strong evidence of validity: The CAPS has excellent convergent and discriminant validity, diagnostic utility, and sensitivity to clinical change. Finally, we address several concerns about the CAPS and offer recommendations for optimizing the CAPS for various clinical research applications.

The Clinician-Administered PTSD Scale: A Review of the First Ten Years of Research

Since its development in 1990 at the National Center for Posttraumatic Stress Disorder (PTSD), The Clinician-Administered PTSD Scale (CAPS; Blake et al., 1990) has become one of the most widely used structured interviews for diagnosing and measuring the severity of PTSD. Initially validated on combat veterans, the CAPS has now been used successfully in a wide variety of trauma populations, including victims of rape, crime, motor vehicle accidents, incest, the Holocaust, torture, and cancer. It has served as the primary diagnostic or outcome measure in more than 200 empirical studies on PTSD and has been translated into at least ten languages. In addition, a child and adolescent version of the CAPS has been developed and is now undergoing field testing and psychometric evaluation. Originally based on the PTSD criteria in the DSM-III-R, the CAPS has been revised several times in response to user feedback and changes in the PTSD diagnostic criteria, with the most significant revision occurring after the publication of the DSM-IV in 1994.

The present paper is an update on the CAPS and a critical review of the first ten years of CAPS-related research. It was prompted by the increasing popularity of the CAPS, the rapid accumulation of empirical evidence supporting its use, and the need to inform current and potential CAPS users about the latest revisions and recommendations for administration and scoring. This paper consists of three sections. First, we provide a brief overview of the CAPS, describing the rationale for its development, its key features, and its evolution through an extensive revision for DSM-IV, as well as a description of other minor modifications. Second, we review the published literature on the CAPS, focusing in particular on psychometric studies of the CAPS and on pharmacological and psychosocial treatment studies that employed the CAPS as an outcome measure. Third, we discuss the implications of the findings and offer recommendations for using the CAPS in a range of research and clinical applications.

This paper was not intended as an in-depth critique of the methodology or conceptual implications of the studies we reviewed, nor did we seek to reach any general conclusions about the current status of PTSD research. Rather, our main purpose was simply to identify studies that have used the CAPS and summarize the empirical findings that bear directly on its psychometric properties and utility for assessing PTSD. Finally, since the child and adolescent version is still undergoing validation, we focus here only on research examining the adult CAPS.

Overview of the CAPS

In developing the CAPS the primary goal was to create a comprehensive, psychometrically sound interview-based rating scale that would be widely accepted as a standard criterion measure of PTSD. In this sense it was intended to serve a role in the field of traumatic stress analogous to that of the ubiquitous Hamilton Depression Rating Scale (HAM-D; Hamilton, 1960) in the field of depression. The CAPS was designed with a number of features intended to improve upon existing PTSD interviews and enhance the reliability and validity of PTSD assessment (see Blake et al., 1995, for a full discussion and a comparison of the CAPS with other PTSD interviews). First, the CAPS can be used either as a dichotomous (present/absent) diagnostic measure or as a continuous measure of PTSD symptom severity. Second, the CAPS assesses both the frequency and intensity of individual PTSD symptoms on separate 5-point (0-4) rating scales, and these ratings can be summed to create a 9-point (0-8) severity score for each symptom. This permits considerable flexibility in scoring: CAPS users can focus on the frequency, intensity, or severity ratings for individual PTSD symptoms, for the three PTSD symptom clusters (reexperiencing, avoidance and numbing, and hyperarousal), and for the PTSD syndrome as a whole.

Third, the CAPS promotes uniform administration and scoring through carefully phrased prompt questions and explicit rating scale anchors with clear behavioral referents. Initial prompt questions explicitly target each symptom, and follow-up prompts help interviewers clarify the inquiry as needed, anticipating typical points of ambiguity or confusion regarding the PTSD criteria. These features

enhance standardization across interviewers and ensure comparability of scores across diverse settings, raters, and trauma populations. Fourth, the CAPS provides complete coverage of the PTSD syndrome. The original version of the CAPS included 17 items assessing the DSM-III-R symptoms of PTSD, 8 items assessing associated features (e.g., guilt, hopelessness, memory impairment), and 5 items assessing response validity, global severity, global improvement, and social and occupational impairment. As described below, the current version of the CAPS assesses all DSM-IV diagnostic criteria for PTSD, including Criterion A (exposure to a traumatic event), Criteria B-D (core symptom clusters of reexperiencing, numbing and avoidance, and hyperarousal), Criterion E (chronology), and Criterion F (functional impairment), as well as the associated symptoms of guilt and dissociation. Finally, the CAPS assesses current and lifetime PTSD symptom status. The prompts for lifetime diagnosis help the interviewer establish explicitly that any endorsed symptoms occurred as a syndrome within the same one-month period.

Initially it was decided that two parallel versions of the CAPS were needed in order to address two distinct assessment needs. The CAPS-1, or current and lifetime diagnostic version, was designed to assess PTSD symptom severity and diagnostic status over the past month, or for the worst month since the trauma. The CAPS-2, or one-week symptom status version, was designed to measure PTSD symptom severity over the past week, and was intended primarily for repeated assessment over relatively brief time intervals in pharmacological research. Apart from the different time frames assessed, the main difference between the CAPS-1 and CAPS-2 is that for the ten CAPS items where symptom frequency is rated in terms of a count (i.e., how often) as opposed to a percentage (i.e., how much of the time), the rating scale anchors on the CAPS-2 were based on a one-week time frame, whereas for the CAPS-1 they were based on a one-month time frame. The distinction between these two original versions of the CAPS led to some confusion in the field, such that the CAPS-2 was thought by some to be a revised version of the CAPS. In response to this confusion, as part of the DSM-IV revision, the CAPS-1 was renamed the CAPS-DX (i.e., CAPS-Diagnostic version), and the CAPS-2 was renamed the CAPS-SX (i.e., CAPS-Symptom Status version). As discussed below, these two versions were recently combined into a single instrument now simply known as the CAPS.

Following the publication of the DSM-IV in 1994, the CAPS was revised, both to bring it up to date with changes in the PTSD criteria and to incorporate user feedback accumulated since its release in 1990. The overarching goal for the revision was to ensure backward compatibility with the original CAPS. This was accomplished by retaining the basic structure, most of the prompt questions, and the values and stems for the rating scale anchors. The revision included four major modifications and a number of relatively minor ones. Major modifications included:

1. Adding a brief protocol for assessing Criterion A (exposure to a traumatic event). This consists of a 17-item self-report checklist of potentially traumatic events and follow-up questions to help the interviewer determine if a stressful event satisfies both parts of the DSM-IV definition of a traumatic event (i.e., the event involves life threat, serious injury, or threat to physical integrity; and the person responds with intense fear, helplessness, or horror).

2. Rewording some of the descriptors for the intensity rating scale anchors. This was done to achieve a consistent focus across items on the three key dimensions of intensity (duration, subjective distress, and functional impairment), to achieve roughly equal gradations of intensity between each of the rating scale values, and to provide examples applicable to a range of trauma populations.

3. Adding a three-point rating scale ("definite," "probable," and "unlikely") that requires interviewers to determine if a reported symptom is attributable to a specific traumatic event. This scale only applies to the last 9 of the 17 symptoms of PTSD (emotional numbing and hyperarousal), because the first 8 symptoms (reexperiencing, effortful avoidance, and amnesia) are all inherently trauma-linked.

4. Replacing six of the eight original associated features. The two items assessing guilt were

retained, but the other items were felt to be either too population-specific (e.g., homicidality, disillusionment with authority) or too broad or complex to be assessed with a single item (e.g., sadness and depression). Also, feedback indicated that they were not routinely administered in most settings. They were replaced with three items assessing the dissociative symptoms of acute stress disorder: reduction in awareness, derealization, and depersonalization. The addition of these items meant that the CAPS could be used to assess acute stress disorder, either currently, if administered within one month of the trauma, or retrospectively.

The minor modifications included: (a) reordering the items to correspond to the order of the DSM-IV diagnostic criteria; (b) adding items to fully assess Criterion E (duration requirement) and Criterion F (subjective distress and functional impairment requirement); (c) renaming the CAPS-1 and CAPS-2, as described earlier; (d) improving the formatting and typeface conventions; (e) eliminating the "at its/their worst" convention for the intensity prompts; (f) eliminating the phrase "without being exposed to something that reminded you of the event" from the frequency prompt for the first item assessing intrusive recollections; and (g) adding an instruction to the interviewer to specify the basis of any QV (questionable validity) ratings.

Completing this discussion on the development of the CAPS are two significant, quite recent developments. One development is the decision to eliminate the "two CAPS" system (i.e., the distinction between the CAPS-1 or CAPS-DX and the CAPS-2 or CAPS-SX) and create a single CAPS scale that can be used to assess PTSD symptoms over the past week, past month, or worst month since the trauma. As noted earlier, the CAPS-2 or CAPS-SX was designed to monitor changes in symptom status over a one-week time frame, and it appears to work well for this purpose, demonstrating excellent psychometric properties (Nagy et al., 1999). The problem, however, is that for the ten CAPS items where symptom frequency is measured as "how often" versus "how much of the time" (i.e., as the number of occurrences rather than as a percentage of time) the CAPS-SX and CAPS-DX had different values because of the different time frame (i.e., for the past week time frame on the CAPS-SX 0=never, 1=once, 2=two or three times, 3=four or five times, 4=daily or almost every day, but for the past month time frame on the CAPS-DX 0=never, 1=once or twice, 2=once or twice a week, 3=several times a week, 4=daily or almost every day).

This means that scores on the two versions were not directly comparable, with CAPS-SX scores tending to yield lower scores when the reported frequency is in the 3-5 times a week range. As a result, investigators who wanted to use the CAPS to establish a PTSD diagnosis as an inclusion criterion, but were interested in weekly assessment intervals over the course of the study, needed to administer a CAPS-DX in the initial evaluation, then administer a CAPS-SX at baseline, mid-treatment, and post-treatment, and then a CAPS-DX at long-term follow-up if they wished to assess end-point diagnostic status. In general this is a workable scheme, but proved to be needlessly cumbersome. Therefore, on the recommendation of the CAPS Advisory Group for the National Center for PTSD, the CAPS-DX and CAPS-SX were combined into a single version, now simply known as the CAPS. This was accomplished by two minor modifications to the CAPS-DX. First, the word "week" was provided as an alternative to "month" in the prompt questions for frequency (e.g., "How often have you had these memories in the past month [*week*]?"). Second, for each item a space was provided to record frequency and intensity ratings for "past week," in addition to "past month" and "lifetime." When the new combined version of the CAPS is used to assess one-week symptom status, frequency ratings for the ten items for which frequency is rated as a count are scored as 0=never, 2=once or twice a week, 3=several times a week, and 4=daily or almost every day, skipping the value 1=once or twice (a month). Thus, the combined CAPS is appropriate for assessing one-month or one-week intervals and yields comparable scores from either application.

The second development involved new options for interpreting CAPS scores. First, nine scoring rules for deriving a PTSD diagnosis have been developed and compared on their psychometric properties and utility for different assessment tasks (Weathers et al., 1999b). It should be emphasized that although

several of these rules appear to be quite useful, more research is needed before firm recommendations can be made. A number of other rules are possible and may prove to have greater utility for some applications. Second, five rationally derived severity score ranges for interpreting CAPS total severity scores have been proposed and are currently being evaluated. These categories are 0-19=asymptomatic/few symptoms, 20-39=mild PTSD/subthreshold, 40-59=moderate PTSD/threshold, 60-79=severe PTSD symptomatology, ≥80=extreme PTSD symptomatology. Finally, a rationally derived 15-point change in CAPS total severity score has been proposed as a marker of clinically significant change. Again, it should be emphasized that these severity score ranges and the 15-point marker are preliminary, and unlike the scoring rules have not been empirically evaluated, but they offer some guidance to clinicians and investigators who use the CAPS to measure change.

In summary, the format and the procedures for administering and scoring the CAPS have evolved in the ten years since it was first developed. However, the changes can be characterized as refinements rather than major revisions, and the goal of backward compatibility of the latest CAPS with the original version appears to have been accomplished (Weathers et al., 1999b). The CAPS now provides a range of options regarding administration and scoring. Interviewers can administer only the 17 core symptoms, all DSM-IV criteria (A-F), or add the associated symptoms. Current symptom status can be assessed for the past week or past month, and lifetime status can be assessed for the worst month since the trauma. By administering the 17 core symptoms plus the 3 dissociative items the CAPS can also be used to assess acute stress disorder. In terms of scoring options, the CAPS can be used to derive a PTSD diagnosis, by using one or more of the available scoring rules, or a continuous severity score for each item, for the three symptom clusters, or for the entire syndrome. Total severity scores summed over the 17 core symptoms can be interpreted with respect to the five proposed severity score ranges, from asymptomatic to extreme, and a 15-point change in CAPS scores can be used to indicate clinically significant change.

Review of the CAPS-Related Literature

Literature Search and Selection of Studies

We developed an initial list of studies to be included by searching the phrase "Clinician-Administered PTSD Scale" in the "Instruments" index of the PILOTS database. PILOTS is the most comprehensive database for the field of traumatic stress, containing virtually every relevant citation in journals and book chapters. This search, conducted in October 1999, yielded 241 citations. We excluded book chapters, review papers, dissertations, letters to the editor, an article on the child and adolescent version of the CAPS, and several studies in which CAPS-related data were included, but not in a form suitable for our purpose. This narrowed the list to a total of 210 studies deemed eligible for potential inclusion in our review.

For the purposes of this review, we divided the eligible studies into three categories: (a) psychometric studies, which provided direct evidence of the reliability and validity of the CAPS; (b) pharmacotherapy and psychotherapy studies, which provided evidence of the sensitivity of the CAPS to clinical change; and (c) case-control studies, which provided additional validity evidence based on conceptually meaningful differences between individuals diagnosed with and without PTSD using the CAPS. In the following sections we summarize all of the available studies in the first two categories since there was a manageable number of them and they provided the richest information regarding the utility of the CAPS. However, due to space constraints, we limit our discussion of studies in the third category to several representative examples, since these were more numerous and provided more limited validity evidence.

As noted earlier, the purpose of this review was to examine all available research addressing the psychometric characteristics of the CAPS and its usefulness as a standard criterion measure of PTSD. Accordingly, we placed few restrictions in selecting the studies to be included, realizing that the final set of studies would vary widely in their quality of design and interpretability of results. We felt that a

consistent pattern of positive results across a large number of studies would provide unambiguous support for the CAPS, and that if the studies varied in quality it would make an even stronger case regarding the generalizability of the findings. In the process of evaluating a psychological assessment instrument, each study, regardless of how well-designed and executed it is, only contributes one piece of evidence and can never be considered definitive. Conclusive answers can be reached only by considering the accumulation of several different types of evidence across different trauma populations, settings, and research designs. In the next section we briefly review some fundamental psychometric concepts in order to provide a conceptual framework for organizing and evaluating the evidence regarding the effectiveness of the CAPS.

Psychometric Considerations

Psychological assessment instruments are evaluated with respect to two important characteristics, reliability and validity. Reliability refers to the consistency of test scores over repeated observations. Three commonly reported types of reliability include internal consistency, test-retest reliability, and interrater reliability, each of which addresses a different potential source of error in test scores. Internal consistency refers to consistency over different items on a test. Requiring only a single administration of a test, it is usually indexed by coefficient alpha (Cronbach's alpha), which ranges from 0.00 to 1.00, with higher values reflecting a greater degree of intercorrelation among the items. Item-scale total correlations, which reflect how well each item correlates with the remaining items, are another useful source of information about internal consistency. Test-retest reliability refers to consistency of test scores over repeated administrations. It is estimated by administering a test twice and calculating the correlation between the two scores. Interrater reliability refers to consistency of test scores over different raters. It is estimated by having two or more raters evaluate and score responses, then calculating either a correlation (only two raters) or intraclass correlation (more than two raters) on the scores. When an instrument is used to obtain a dichomotomous score, as in the case of a present/absent diagnostic decision, interrater or test-retest reliability is estimated by calculating a kappa coefficient, a chance-corrected measure of agreement.

Two different research designs are typically employed to evaluate the reliability of a structured interview such as the CAPS. In a simple interrater design, two or more raters independently rate the same interview. One rater administers and scores the interviews as usual, while additional raters either observe the interview live or, if more convenient, observe an audiotape or videotape of the interview. Since the information available to the raters is identical, the only potential source of error is inconsistency in scoring among raters. In a test-retest design, two independent raters administer and score the interview on separate occasions. This is a more stringent test of reliability because it involves inconsistency in scoring plus two additional potential sources of error: inconsistency in how raters ask the questions and inconsistency in respondents' answers. Although we follow common practice in referring to it as test-retest reliability, the reliability estimate this design yields is more precisely known as a coefficient of stability and interrater equivalence, because it involves both occasions and raters as potential sources of error. An important consideration for the test-retest design is the interval between interviews. If the interval is too brief, respondents' answers in the second interview may be influenced by their memory of their answers in the first interview. If the interval is too long, genuine change in clinical status may occur, meaning that inconsistencies in responses are legitimate and not a source of error. In the assessment of PTSD, an interval of a few days to a week is probably reasonable for most applications.

Although reliability clearly is a desirable characteristic of an assessment instrument, a more important concern is validity, which refers to the extent to which evidence exists to support the various inferences, interpretations, conclusions, or decisions that will be made on the basis of a test. Traditionally, three types of validity have been identified. The first type is content validity, which refers to evidence that items on a test adequately reflect the construct being assessed. The second type is criterion-related validity, which refers to evidence that the test can predict some variable or criterion of interest. The criterion may be measured either at the same time the test is administered (concurrent

validity) or at some point after the test (predictive validity). The third type is construct validity, which refers to evidence that the test measures the construct of interest and not other constructs. This can be demonstrated, for example, by showing that the test correlates strongly with other measures of the same construct (convergent validity) but not with measures of other constructs (discriminant validity).

However, this traditional approach to validity has recently been superseded by the latest revision of the *Standards for Educational and Psychological Testing* (*Standards*; 1999), which maintains:

> "[Different] sources of evidence may illuminate different aspects of validity, but they do not represent distinct types of validity. Validity is a unitary concept. It is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose. Like the 1985 *Standards*, this edition refers to types of validity evidence, rather than distinct types of validity." (p. 11)

Thus, the new *Standards* argues for an integrative approach to validity, emphasizing a confluence of validity evidence from different sources, and its updated scheme for categorizing validity evidence represents a marked departure from previous editions. Categories include: (a) evidence based on test content; (b) evidence based on response processes, which focuses on respondents' behavior during the test process; (c) evidence based on internal structure, which focuses on relationships among test items and components; (d) evidence based on relations to other variables, which includes convergent and discriminant evidence, criterion-related evidence, and the generalization of validity to new testing situations; and (e) evidence based on consequences of testing, which focuses on both the intended and unintended outcomes of test use.

The new *Standards* also emphasizes that the process of validation applies not to tests themselves, but rather to any specific interpretations that will be made on the basis of test scores. Therefore, stating that a test is valid begs the question: Valid for what purpose? To address this question specifically with regard to the CAPS, two main uses of the CAPS have been proposed. One is to establish a dichotomous PTSD diagnosis and the other is to provide a continuous measure of PTSD symptom severity. Thus, the two main interpretations of CAPS scores that should be the focus of validation are:

1. CAPS scores reflect severity of PTSD symptoms, for individual symptoms, symptom clusters, or the syndrome as a whole.

2. CAPS diagnoses reflect the presence or absence of PTSD.

One source of validity evidence that applies to these inferences is content-based evidence. This refers to the extent to which the content of a test corresponds to the construct being assessed. In this regard, the CAPS was written and revised by a team of experts in traumatic stress at the various branches of the National Center for PTSD. It was based directly on the diagnostic criteria for PTSD in the DSM-III-R, and now DSM-IV, and represents these criteria faithfully. As noted earlier, the major revision of the CAPS that followed the publication of the DSM-IV not only reflected changes in the PTSD criteria, but also took into account formal and informal feedback from a broad cross-section of CAPS users in other clinical research settings. Although difficult to quantify, there is clearly a consensus among those familiar with the CAPS that the content of the CAPS corresponds veridically to the construct of PTSD.

A second source of validity evidence has to do with the internal structure of the CAPS. As currently conceptualized in the DSM-IV criteria, PTSD is a multifaceted syndrome consisting of three closely related but distinct symptom clusters, reexperiencing, avoidance and numbing, and hyperarousal. If PTSD is a syndrome, then there should a reasonably high degree of correlation among all of the symptoms. If there are distinct but overlapping symptom clusters, then the items within the clusters should correlate more strongly with each other than they do with symptoms in other clusters. These relationships would be reflected in alpha coefficients and item-total correlations. Factor analysis, especially confirmatory factor analysis, in which competing hypotheses about the nature of PTSD can be directly compared, is another means of evaluating the internal structure of the CAPS.

A third, and particularly important, source of validity evidence involves the relationship between

the CAPS and other variables. As conceptualized in the latest Standards, this source of evidence includes what used to be referred to as construct and criterion-related validity, and encompasses a broad range of evidence that the CAPS corresponds in theoretically meaningful ways with measures of other constructs. Relevant findings might include: (a) convergent evidence, showing relatively strong correlations between the CAPS and other measures of PTSD; (b) discriminant evidence, showing relatively weak correlations between the CAPS and measures of different constructs; (c) evidence of test-criterion relationships, showing the correspondence between the CAPS and a criterion such as a PTSD diagnosis or an indicator of clinically significant improvement in PTSD symptom severity; (d) evidence that groups formed on the basis of the CAPS differ as hypothesized on some characteristic or behavior; and (e) evidence that PTSD prevalence, severity, or symptom profile based on the CAPS vary as hypothesized in different groups.

Psychometric Studies

In this section, we describe the results of studies that emphasized the psychometric properties of the CAPS, including studies in which the CAPS was either the primary instrument being investigated or was included as a validational measure for another PTSD instrument. First, we summarize studies that examined reliability and convergent and discriminant validity. Then we summarize studies that address two other psychometric issues: the factor structure of the CAPS and the utility of various scoring rules for converting CAPS frequency and intensity scores into a dichotomous PTSD diagnosis. In reviewing these studies, we found that investigators often neglected to specify the version of the CAPS they administered and the scoring rule they used to determine a PTSD diagnosis. The version could usually be readily inferred, and with few exceptions was the CAPS-1 or CAPS-DX, the current and lifetime diagnostic version. In our discussion of the studies in this section, then, "CAPS" refers to the CAPS-1 or CAPS-DX, and "CAPS-2" is used explicitly to refer to the weekly symptom-rating version. Unless explicitly stated, however, the scoring rule could not be determined. For the purposes of this review we assumed, unless stated otherwise, that investigators used the original scoring rule, whereby a frequency of "1" or higher and an intensity of "2" or higher for a given CAPS item indicated symptom endorsement.

Reliability, Convergent and Discriminant Validity, and Diagnostic Utility

The CAPS has been the primary focus of several psychometric investigations. Blake et al. (1990) reported the first psychometric data on the CAPS. In a pilot study they administered the CAPS, the Combat Exposure Scale (CES; Keane et al., 1989), the Mississippi Scale for Combat-Related PTSD (Mississippi Scale; Keane et al., 1988), and the Keane PTSD Scale of the MMPI (PK scale; Keane et al., 1984) to 25 male combat veterans. To determine interrater reliability for the CAPS, a second rater observed and independently rated seven interviews. Excellent agreement was found between the two raters, with reliability coefficients for frequency and intensity scores across the three symptom clusters (reexperiencing, numbing and avoidance, and hyperarousal) ranging from .92 to .99. The raters also demonstrated perfect diagnostic agreement for the seven participants, five of whom had a positive diagnosis. Internal consistency for the three PTSD symptom clusters was high, with alpha coefficients ranging from .73 to .85 for the three symptom clusters. Regarding convergent validity, the CAPS correlated strongly with the Mississippi Scale (.70) and the PK scale (.84). It also correlated .42 with the CES, a moderate correlation that is typical for correlations between measures of trauma exposure and measures of PTSD.

Hovens et al. (1994) examined the psychometric properties of the CAPS in a Dutch sample, employing translations of the CAPS and other PTSD measures. Participants were 76 Dutch trauma survivors (51 males, 25 females), including combat veterans, resistance veterans, and concentration camp survivors. Participants were first diagnosed with or without PTSD, using DSM-III-R criteria, on the basis of an unstructured clinical interview. They were then administered the CAPS, the Mississippi Scale, the PK scale, and the IES. Interrater reliability on the CAPS was evaluated through simultaneous ratings of nine interviews by two independent clinicians. Diagnostic agreement was perfect for these nine participants. Further, reliability coefficients for frequency and intensity scores for individual items were

strong, ranging from .59 to 1.00 for frequency, with a mean of .92, and .52 to 1.00 for intensity, with a mean of .86. At the symptom cluster level, reliability coefficients ranged from .92 to 1.00 for frequency and .92 to .98 for intensity. Regarding internal consistency, Hovens et al. found alphas of .63 for reexperiencing, .78 for avoidance and numbing, .79 for hyperarousal, and .89 for all 17core PTSD symptoms. No rationale was given for the decision to report internal consistency for intensity scores but not for frequency or severity (frequency + intensity) scores.

Using the clinical interview as the criterion, Hovens et al. found that a CAPS-based PTSD diagnosis had 74% sensitivity, 84% specificity, and 79% efficiency, and a kappa of .58. Because these figures were lower than expected, they examined discrepancies between the clinical interview and the CAPS. They concluded that in the clinical interview clinicians primarily emphasized reexperiencing symptoms in making a PTSD diagnosis, failing to give sufficient attention to the other two symptom clusters, particularly avoidance and numbing. They further found that many of the participants with discrepant diagnoses were only mildly symptomatic, and thus more diagnostically ambiguous. As evidence of convergent validity, the total CAPS score correlated .73 with the Mississippi Scale, .74 with the PK scale, and .62 with the IES total score. Finally, with the exception of amnesia, the prevalence of each of the 17 core PTSD symptoms on the CAPS was significantly greater in participants with PTSD than in those without PTSD, indicating robust discrimination between the two groups.

As part of an effort to develop and evaluate a computer-administered version of the CAPS, Neal et al. (1994) administered both the computerized and the original interview versions of the CAPS to 40 military personnel (36 males, 4 females) with mixed trauma exposure, including combat, non-combat-related assaults, accidents, and disasters, and childhood physical and sexual abuse. To evaluate the reliability of the CAPS interview, ten participants were interviewed twice by independent clinicians, resulting in perfect diagnostic agreement. Treating the CAPS interview as the criterion, the computerized version had 95% sensitivity and 95% specificity, with a kappa of .90. Although the interval between the two versions was not specified, they appear to have been administered in a single session, which could have inflated this high level of agreement. An initial finding of a high correlation (.96) between total frequency and total intensity scores on both the interview and computerized versions of the CAPS led Neal et al. to use intensity scores alone as a continuous measure of severity in all further analyses. Internal consistency of intensity scores was high for both versions, with an alpha of .90 and a median item-total correlation of .77 for the interview version, and an alpha of .92 and a median item-total correlation of .70 for the computerized version. In addition, intensity scores on the two versions were strongly correlated, ranging from .55 to .92 for individual items and from .87 to .92 for the three symptom clusters. The correlation for total intensity score between the two versions was .95.

Hyer et al. (1996) investigated the utility of the CAPS for assessing older combat veterans. Participants were 125 male World War II and Korean combat veterans. They were administered a computer-assisted version of the SCID (SCID-DTREE), including the PTSD module, as well as the CAPS, by two clinicians. They also completed the Mississippi Scale, the IES, and the CES. To assure the comparability of the SCID-DTREE and the SCID, 25 participants were administered the SCID in a separate testing session by an independent clinician. In this subsample there was perfect agreement as to PTSD diagnostic status, not only between the SCID-DTREE and the SCID, but between the CAPS and the SCID. In the full sample, against a PTSD diagnosis based on the SCID-DTREE, the CAPS had 90% sensitivity, 95% specificity, and 93% efficiency, and a kappa of .75. The CAPS also demonstrated high internal consistency, with alphas of .88 for reexperiencing, .87 for avoidance and numbing, .88 for hyperarousal, and .95 for all 17 core items. CAPS diagnosis was correlated .81 with the IES, .61 with the Mississippi Scale, and .26 with the CES. The relatively low correlation with the CES is likely attributable in part to a restricted range on the CES, since most participants had moderate to heavy combat exposure.

As part of a large prospective study on the effects of trauma, Shalev et al. (1997) employed signal detection methodology to determine whether the CAPS or any of several questionnaire measures of PTSD, dissociation, and anxiety administered at one week or one month post-trauma could predict PTSD

diagnostic status at four months post-trauma. Participants included 207 (98 male, 109 female) victims of civilian trauma recruited from the emergency room of a hospital. In most cases, the traumatic event involved a motor vehicle accident. Within a week of their trauma, participants completed the IES, the State form of the State Trait Anxiety Inventory (STAI; Spielberger et al., 1970), and the Peritraumatic Dissociative Experiences Questionnaire (PDEQ; Marmar et al., 1997). Assessments at one month and four months post-trauma added the CAPS and the civilian version of the Mississippi Scale to this battery. They found that all of the questionnaires administered at either one week or one month post-trauma were predictive of PTSD diagnostic status at four months, but that none of the questionnaires differed significantly in terms of accuracy of prediction. In contrast, the CAPS at one month post-trauma, used as a continuous measure, was a significantly better than all of the questionnaires in predicting a four-month diagnostic status that was also based on the CAPS. Although Shalev et al. did not identify an optimal cutoff score for CAPS total severity, they did provide diagnostic utility data for a range of selected cutoff scores. These data indicate that a CAPS score of 40 yielded 93% sensitivity and 80% specificity.

To determine the prevalence of PTSD in veterans with spinal cord injuries, Radnitz et al. (1995) administered the CAPS and the SCID PTSD module to 126 male veterans receiving medical care for spinal cord injuries in inpatient and outpatient settings. Current and lifetime diagnostic status was assessed on both the CAPS and the SCID. To determine diagnostic status on the CAPS, Radnitz et al. used a variant of the original scoring rule (i.e., frequency $\geq$ 1, intensity $\geq$ 2), whereby either the frequency or intensity of an item had to be "2" or higher and the other dimension had to be a "1" or higher. As described below, this scoring rule was referred by Blanchard et al. (1995) as the "Rule of 3." Although Radnitz et al. did not provide kappas or other diagnostic utility statistics except efficiency, we were able to calculate these from data provided in the tables. Treating the SCID as the criterion, for current diagnosis the CAPS had 83% sensitivity, 94% specificity, 93% efficiency, and a kappa of .73. For lifetime diagnosis the CAPS had 84% sensitivity, 90% specificity, 88% efficiency, and a kappa of .74. Although not explicitly stated, it appears that both interviews were administered by the same research assistant in the same session. Both of these factors, i.e., the lack of a time interval between interviews and the lack of an independent rater, could have inflated the correlation between the CAPS and the SCID. Finally, CAPS total severity scores appeared to strongly differentiate between participants with and without a PTSD diagnosis, although these mean differences were not evaluated by statistical test.

Although all of these studies provide valuable information, the most comprehensive investigations of the psychometric properties of the CAPS, based on data collected at the National Center for PTSD, are described in two articles currently submitted for publication. Weathers et al. (1999a) examined the reliability and validity of the CAPS-1 / CAPS-DX in five samples of male Vietnam veterans, including 267 veterans from four different research projects and 571 veterans seen for clinical services. To evaluate the test-retest reliability (i.e., stability and rater equivalence) of the CAPS-1, 60 veterans were administered the CAPS twice, at a 2-3 day interval, by independent clinicians. For the three symptom clusters intraclass correlations ranged from .86 to .87 for frequency, .86 to .92 for intensity, and .88 to .91 for severity. Across all 17 symptoms intraclass correlations were .93 for total frequency, .95 for total intensity, and .95 for total severity. Following the revision of the CAPS for DSM-IV, the same design was implemented for the CAPS-DX in a smaller sample of 24 veterans. This study also yielded robust estimates of reliability, with intraclass correlations of .91 for total frequency, .91 for total intensity, and .92 for total severity. Using the optimal scoring rule, kappa, indicating test-retest reliability for a CAPS-based PTSD diagnosis, was .89 in the first sample and 1.00 in the second sample.

Examining internal consistency, Weathers et al., in a combined research sample of 243 veterans, found alphas for the three symptom clusters ranging from .78 to .87 for frequency, .82 to .88 for intensity, and .82 to .88 for severity. Alphas for all 17 items were .93 for frequency, .94 for intensity, and .94 for severity. In the clinical sample, alphas for the three symptom clusters ranged from .64 to .73 for frequency, .66 to .76 for intensity, and .69 to .78 for severity. Alphas for all 17 items were .85 for frequency, .86 for intensity, and .87 for severity. The lower alphas in the clinical sample were likely due

in part to a restricted range in CAPS scores, since most veterans referred for clinical services at the National Center report moderate to severe PTSD symptoms; they may also be due to a much larger and more diverse pool of clinicians, relative to the small number of well-calibrated clinicians who administered the CAPS to the research samples. Nonetheless, these scores provide excellent evidence supporting the CAPS as used in a clinical setting.

Weathers et al. also reported validity evidence for the CAPS, focusing primarily on convergent and discriminant validity evidence and the diagnostic utility of the CAPS against a PTSD diagnosis based on the SCID. In the first research sample of 123 veterans, the CAPS total severity score correlated .53 with the CES, .91 with the Mississippi Scale, .77 with the PK scale, .89 with the number of PTSD symptoms endorsed on the SCID, and .94 with the PTSD Checklist (PCL; Weathers et al., 1993), a 17-item self-report measure of PTSD. CAPS total severity correlated somewhat less strongly, but still robustly, with measures of depression (.61 to .75) and anxiety (.66 to .76), findings that were expected given the substantial overlap between PTSD, depression, and anxiety. Much weaker correlations were observed between CAPS total severity and measures of antisocial personality (.14 to .33), a disorder conceptually distinct from PTSD. In an effort to bring these convergent and discriminant correlations into sharper relief, Weathers et al. then calculated partial correlations, controlling first for nonspecific distress and symptom exaggeration by using the F scale of the MMPI-2, then for nonspecific distress again using the Global Severity Index (GSI) of the Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1983). After controlling for the F scale, the CAPS demonstrated strong partial correlations with measures of PTSD, including the Mississippi Scale (.83), the PCL (.89), and the number of PTSD symptoms on the SCID (.82). As predicted, however, partial correlations between the CAPS and measures of depression (.37 to .53) and anxiety (.37 to .55) were markedly lower, and those between the CAPS and measures of antisocial personality were essentially zero (-.05 to .02). A similar, but even more striking pattern was found after controlling for the GSI. Further, these results involving the F scale were generally replicated in a second research sample.

Finally, again focusing on the sample of 123 participants, Weathers et al. reported the diagnostic utility of three CAPS scoring rules for predicting a SCID-based PTSD diagnosis. The original, rationally derived scoring rule (frequency $\geq$ 1, intensity $\geq$ 2, or F1/I2) had 91% sensitivity of, 71% specificity, and 82% efficiency, with a kappa of .63. These figures reveal the F1/I2 rule to be relatively lenient, with excellent sensitivity but only moderate specificity, suggesting that it tends to somewhat overdiagnose PTSD relative to the SCID. The two other rules were empirically derived on this sample. The second rule, which assigns a positive diagnosis if the CAPS total severity score is 65 or greater (TSEV65), had 82% sensitivity, 91% specificity, and 86% efficiency, with a kappa of .72. Although the higher kappa indicates a better correspondence with the SCID than the F1/I2 rule has, the TSEV65 rule appears to be relatively stringent, tending to somewhat underdiagnose PTSD relative to the SCID. The third rule, derived by empirically calibrating each CAPS symptom with the analogous SCID symptom (SXCAL), had the closest correspondence to the SCID, with 91% sensitivity, 84% specificity, and 88% efficiency, with a kappa of .75. Although these results require cross-validation, the SXCAL rule appears to be the optimally efficient rule, and therefore the best choice for differential diagnosis.

In the second article based on National Center data, Nagy et al. (1999) described the only comprehensive investigation of the CAPS that focused specifically on the CAPS-2. To evaluate interrater reliability, Nagy et al. administered the CAPS-2 to 30 (29 male, 1 female) inpatients and outpatients in treatment for PTSD, all but two of whom were combat veterans. Interviews were videotaped and scored by three additional raters, resulting in four ratings for each participant. Intraclass correlations ranged from .76 to .99 for the 17 core PTSD symptoms, and from .92 to .97 for the three symptom clusters, with values of .98 for total frequency, .96 for total intensity, and .98 for total severity. Internal consistency and convergent and discriminant evidence were examined in two additional samples of male combat veterans: 20 veterans enrolled in a pharmacologic trial and 37 veterans in inpatient PTSD treatment program. All participants were administered the CAPS and the IES. In addition, the 20 participants in the drug trial

were administered the Hamilton scales for depression and anxiety (HAM-D and HAM-A), and the 37 inpatients completed the BDI and the Beck Anxiety Inventory (BAI; Beck et al., 1988).

In the combined sample alphas were .25 for reexperiencing, .69 for avoidance and numbing,.70 for hyperarousal, and .79 for all 17 items. In the combined sample, the CAPS correlated .37 with the IES. For the participants in the drug trial the CAPS correlated .34 with the HAM-D and .36 with the HAM-A. In the inpatient sample the CAPS correlated .67 with the BDI and .51 with the BAI. Taken together, these results are generally in line with results from studies involving the CAPS-1. However, the alpha for the reexperiencing cluster and the correlation of the CAPS with the IES were lower than those found previously. It is unclear whether these findings are sample-specific and reflect some idiosyncrasies of the particular participants or settings in the study, or whether they are attributable to some aspect of the CAPS-2.

Although not designed primarily as psychometric investigations of the CAPS per se, other investigations have nonetheless provided additional evidence of its reliability and validity. Hovens et al. (1994) used the CAPS as a criterion measure in the evaluation of a new self-report measure of PTSD, the Self-Rating Inventory for Posttraumatic Stress Disorder (SIP). The SIP consists of 51 items, 22 assessing DSM-III-R PTSD symptoms and 29 measuring other trauma-related sequelae, particularly those associated with the proposed diagnostic category of disorders of extreme stress not otherwise specified (DESNOS). This study included two samples: the same 76 participants used in their previous study on the CAPS, plus 59 (22 male, 37 female) psychiatric outpatients. Although the psychiatric outpatients were not selected on the basis of a known trauma history, 18 of them reported exposure to various types of civilian trauma, including sexual and physical assault, traumatic loss of a loved one, and motor vehicle accidents. Combining all participants with a trauma history across the two samples, Hovens et al. found that the CAPS correlated .73 with total SIP score, .75 with the DSM-III-R items on the SIP, .70 with the civilian version of the Mississippi Scale, .72 with the PK scale, and .61 with the IES. Correlations for the DSM-III-R symptom clusters between the CAPS and the SIP were .54 for reexperiencing, .69 for avoidance and numbing, and .71 for hyperarousal.

Two studies by Neal and colleagues also provide convergent validity evidence. First, Neal et al. (1994) assessed 70 (59 male, 11 female) military personnel with mixed military and civilian trauma exposure, presumably similar to, or overlapping with, the sample they evaluated for their study on the computerized CAPS described earlier. They examined the correlations of two CAPS variables, total intensity and number of symptoms endorsed, with the PK scale, the IES, and the GSI of the SCL-90. Although no rationale was offered for why they used intensity rather than severity scores, presumably this was because of the high degree of correlation between frequency and intensity scores they found in their previous study on the computerized CAPS. Similar patterns of correlations were found for both CAPS variables. Total CAPS intensity correlated .85 with the PK scale, .78 with the IES, and .77 with the SCL, whereas the number of CAPS symptoms correlated .84 with the PK scale, .81 with the IES, and .74 with the SCL-90. The strong correlations with the PK scale and the IES offer convergent evidence, but the nearly as strong correlations with the SCL-90 failed to provide strong discriminant evidence. However, given the high rates of comorbidity found in PTSD and given that the SCL-90 primarily reflects nonspecific distress, the SCL-90 is not an optimal measure for discriminant evidence. Second, Neal et al. (1995) administered the CAPS, the IES, the PK scale, and the Mississippi scale to 30 (29 male, 1 female) World War II prisoners of war. In this study the CAPS correlated .63 with the IES, .71 with the PK scale, and .81 with the Mississippi Scale, again providing convergent evidence for the CAPS as a measure of PTSD.

Two studies by Blanchard and colleagues provide evidence regarding interrater reliability and convergent validity. Blanchard et al. (1995b) employed the CAPS as the primary diagnostic measure of PTSD in a study of male and female motor vehicle accident victims. All CAPS interviews were audiotaped and an independent rater re-scored 15 randomly selected interviews. Interrater reliability for individual items ranged from .82 to .99, with a mean of .98, and kappa for a PTSD diagnosis was .81.

Blanchard et al. (1996) also used the CAPS as the criterion measure in a psychometric evaluation of the PCL. Participants were 27 (3 male, 24 female) motor vehicle accident victims and 13 female sexual assault victims. Interrater reliability for the CAPS, based on 19 audiotaped and independently re-scored interviews, was again quite strong. Coefficients for individual items ranged from .84 to .99 for individual items, with a mean of .94, and kappa for a PTSD diagnosis was .84. Correlations between the PCL and the CAPS supplied convergent evidence. Correlations between PCL items and corresponding CAPS items ranged from .39 to .79, with all but three correlations above .60 and seven correlations above .70. In addition, the correlation between the total scores on the PCL and the CAPS was .93.

Finally, two studies utilized the CAPS in the validation of the Davidson Trauma Scale (DTS; Davidson et al., 1997), a 17-item self-report measure of DSM-IV PTSD symptoms. Like the CAPS, the DTS assesses PTSD symptoms on two dimensions: frequency, which corresponds to the frequency dimension on the CAPS, and severity, which corresponds to the intensity dimension of the CAPS. The DTS assesses symptoms over the previous week. To obtain convergent evidence for the DTS, Zlotnick et al. (1996) administered the DTS and the CAPS to 50 female sexual abuse survivors. They found correlations of .72 between DTS total frequency and CAPS total frequency and .57 between DTS total severity and CAPS total intensity. For total DTS and total CAPS scores for each of the three symptom clusters, they found correlations of .70 for reexperiencing, .53 for avoidance, and .73 for hyperarousal. As part of a comprehensive psychometric investigation of the DTS, Davidson et al. (1997) administered the DTS and the CAPS to a mixed sample of 102 female sexual assault victims and male combat veterans, finding a correlation of .78 between total scores on the DTS and CAPS.

CAPS Factor Structure

The final two issues we will discuss address the factor structure of the CAPS and the development and evaluation of various scoring rules for deriving a CAPS-based PTSD diagnosis. Two studies have examined the factor structure of the CAPS using confirmatory factor analysis. Buckley et al. (1998) tested a single hypothesized factor structure consisting of two factors: (a) Intrusion and Avoidance and (b) Hyperarousal and Numbing. Although these factors cut across the three DSM-III-R and DSM-IV symptom clusters of PTSD, there is theoretical and empirical justification for this two-factor structure. In fact, Buckley et al. sought to replicate a previous study by Taylor et al. (1998), in which this structure was derived in an exploratory factor analysis. Analyzing CAPS scores from a combined sample of 217 male and female motor vehicle accident victims, Buckley et al. found support for the hypothesized two-factor structure across several indices of model fit.

In a more comprehensive analysis, King, King, Leskin, and Weathers (1998) conducted a confirmatory factor analysis of CAPS scores in 524 male combat veterans seen for clinical services at the National Center for PTSD in Boston. In this study King et al. tested four competing models, three of which involved dividing Criterion C (numbing and avoidance) into two distinct factors of effortful avoidance (criteria C1 and C2) and emotional numbing (criteria C3-C7). The first model was a four-factor, first-order solution consisting of four correlated primary factors: reexperiencing, effortful avoidance, emotional numbing, and hyperarousal. The second model, which was similar to the one Buckley et al. evaluated, was a two-factor, higher order solution, with one factor comprising reexperiencing and effortful avoidance and the other comprising emotional numbing and hyperarousal. The third model was a single factor, higher order solution that hypothesized a single PTSD factor comprising the four symptom clusters. The fourth model was a single-factor, first-order solution that hypothesized that all 17 symptoms load on a single PTSD factor. King et al. found that the first model provided the best fit to the data, suggesting that PTSD, as assessed by the CAPS, consists of four correlated but distinct symptom clusters. This finding supports the CAPS as a measure of PTSD in that the internal structure of the CAPS corresponds to the DSM PTSD symptom clusters, albeit with the additional, conceptually meaningful distinction between effortful avoidance and emotional numbing.

CAPS Scoring Rules

Finally, one of the recent developments in the CAPS has been the explication and evaluation of various rules for converting continuous CAPS scores into a dichotomous PTSD diagnosis. From the outset it was recognized that the original, rationally derived F1/I2 rule described earlier was only an initial working rule that might be replaced by others once sufficient empirical evidence had accumulated. Over time, a number of new rules have been proposed and have recently appeared in the literature. Blanchard et al. (1995a) were the first to compare the impact of adopting different scoring rules. In an investigation of 100 (35 male, 65 female) motor vehicle accident victims they proposed and evaluated three different scoring rules, all of which involved converting CAPS frequency and intensity scores into a dichotomous score for each symptom, then following the DSM requirements (one reexperiencing symptom, three numbing and avoidance symptoms, two hyperarousal symptoms) to derive a PTSD diagnosis. According to the Rule of 2, a symptom is considered present if the severity score for an item (frequency + intensity) is ≥ 2 (i.e., frequency and intensity are both ≥ 1). Similarly, the Rule of 3 requires an item severity score ≥ 3 (either frequency or intensity is ≥ 2 and the other is ≥ 1). This is similar to but more inclusive than the original F1/I2 rule. Last, the Rule of 4 requires an item severity ≥ 4. Blanchard et al. found that the three rules yielded markedly different PTSD prevalence estimates, with 44% for the Rule of 2, 39% for the Rule of 3, and 27% for the Rule of 4. Further, they found that participants who met the Rule of 4 had higher scores on measures of depression and anxiety, and greater functional impairment, relative to those who only met the Rule of 3.

More recently, Weathers et al.(1999b) described and compared nine scoring rules, drawing on data from the same five samples in the Weathers et al. (1999a) psychometric article described earlier. Four of the nine rules were rationally derived, including the original F1/I2 rule, the Item Severity ≥ 4 (ISEV4) rule, which is identical to Blanchard's Rule of 4, and two rules based on clinicians' judgments regarding which frequency/intensity combinations constitute a symptom. The other five rules were empirically derived, including four rules calibrated in various ways against the SCID PTSD module, and one rule identified by Orr (1997), based on a study of physiological reactivity in female incest survivors. Kappa coefficients indicating test-retest reliability for the rules ranged from .72 to .90 in an initial sample of 60 veterans, and from .68 to 1.00 in a followup sample of 24 veterans. Kappa coefficients for predicting a PTSD diagnosis based on the SCID ranged from .63 to .75. As in the Blanchard et al. (1995a) study, the nine rules yielded widely varying prevalence estimates, ranging from 26% to 49% in a combined research sample of 243 veterans and 47% to 82% in a clinical sample of 571 veterans. The F1/I2 rule was the most lenient in the clinical sample and second most lenient in the research sample. The two rules based on clinicians' ratings were the most stringent in both samples. Also, compared to participants who met criteria only by the F1/I2 rule, those who met criteria for the most stringent rule had significantly higher scores on measures self-report measures of PTSD, depression, anxiety, and nonspecific distress.

A third study, by Fleming and Difede (1999), examined the impact of adopting different scoring rules on the CAPS-2 in a sample of hospitalized burn patients. Although they recognized that the CAPS-2 was not suitable for a diagnosis of PTSD because of the one-week time frame, they deliberately chose it for their study because they were interested in acute PTSD symptoms within the first two weeks after the trauma. Administering the CAPS-2 to 69 (48 male, 21 female) participants, they compared the effects of adopting essentially the same scoring rules described by Blanchard et al. (1995a). The one exception was that Fleming and Difede appear to have used the F1/I2 rule rather than Blanchard et al.'s more inclusive Rule of 3. Compared to the previous two studies, they found less variability among the different rules in terms of estimated prevalence of PTSD. The Rule of 3 and the Rule of 4 both yielded a prevalence of 25%, while the Rule of 2 yielded a prevalence of 32%. Further, they found no significant differences on the IES or self-report measures of acute stress and nonspecific distress between participants who met criteria only by the Rule of 2 and those who met criteria by the Rule of 3 or the Rule of 4. However, differences were found on all self-report measures between all participants who met criteria for PTSD by at least the Rule of 2 and those who did not meet criteria for PTSD by any of the rules.

Taken together these three studies of scoring rules for the CAPS indicate that there are important consequences to adopting a particular rule. Prevalence estimates can vary considerably and participants who meet criteria by lenient rules may be less symptomatic and less impaired relative to those who meet criteria by more stringent rules. Weathers et al. (1999b) discuss three implications of these findings. First, investigators should always explicitly describe and defend their choice of a CAPS scoring rule. Second, for many applications, an efficient and informative strategy would be to use several scoring rules, ranging from lenient to stringent, and compare the different results obtained. Third, when using different scoring rules is not feasible, investigators should select scoring rules that are best suited for the purpose of the study. Lenient scoring rules are most appropriate for screening, when a lower threshold for diagnosis is needed to avoid false negatives. Stringent rules are most appropriate for confirming a diagnosis or creating an unambiguous PTSD group for case-control research, when a higher threshold is need to avoid false positives. Moderate rules are most appropriate for differential diagnosis, when false negatives and false positives are weighted equally and the goal is to minimize the overall number of diagnostic errors.

Finally, we note that it is possible that some of the diagnostic utility data cited for the CAPS in this section, even though it is consistently high, might actually have been stronger had different scoring rules been applied. In discussing the articles in this section we assumed that unless stated otherwise, investigators used the original F1/I2 rule to derive a PTSD diagnosis from CAPS scores. However, the Weathers et al. (1999b) article, in particular, demonstrated that the F1/I2 rule is a relatively liberal rule and may not be optimal for differential diagnosis.

Discussion

Considering all the accumulated evidence, the CAPS appears to have excellent psychometric properties across a wide variety of clinical research settings and trauma populations. Interrater reliability for continuous CAPS scores was consistently at the .90 level and above, with diagnostic agreement at times reaching 100%. Test-retest reliability, a more stringent measure of agreement, was nearly as strong, although it was only evaluated in one study and needs replication. These findings suggest that trained and calibrated raters can achieve a high degree of consistency in using the CAPS to diagnose PTSD and rate PTSD symptom severity. In addition, internal consistency was generally high, with alphas typically in the .80 to .90 range for the three PTSD symptom clusters and for the entire syndrome.

Although somewhat more variable and therefore more difficult to easily summarize, evidence of validity was also strong. Regarding convergent evidence, the CAPS generally demonstrated correlations at the .70 level and above with self-report measures of PTSD such as the Mississippi Scale, the PK scale, the IES, the PCL, and the DTS, often reaching the .80 to .90 range. Diagnostic utility of the CAPS was evaluated in five studies, and with one exception, in which the criterion was a clinical diagnosis based on an unstructured interview, was quite robust, with sensitivities and specificities above .80, and often above .90, and kappas above .70. To date, however psychometric studies of the CAPS offer little in terms of discriminant evidence. More data on this are needed. Because individuals with PTSD, especially chronic PTSD, often have comorbid disorders and experience high levels of distress, it may prove to be difficult to obtain unequivocal discriminant evidence, particularly with measures of depression and anxiety, since these two constructs overlap conceptually with PTSD. Weathers et al. (1999a) tried to address this problem by including measures of a construct conceptually unrelated to PTSD (antisocial personality) and by partialing out the effects of nonspecific distress. These two approaches appeared to be successful in providing discriminant evidence, but more creative research on this issue is needed.

We close this section with a brief discussion of some fundamental questions regarding the psychometric investigation of the CAPS. First, regarding convergent and discriminant evidence there are no absolute standards for what constitutes “good” evidence. How large should convergent validity coefficients be? How small should discriminant validity coefficients be? How large of a difference should there be between convergent and discriminant coefficients? Reasonable answers to these questions must be informed by a well-articulated theoretical model and ultimately based on expert

judgment.

Second, is it appropriate to evaluate a putative "gold standard" such as the CAPS against self-report measures? When a correlation between the CAPS and another measure is lower than expected it is unclear if the "problem" lies with the CAPS or with the alternative measure, or a combination of both. This question is particularly important with respect to self-report measures of PTSD, which are subject to misinterpretation and to response biases such as social desirability, exaggeration, minimization, and even random responding. In addition, they vary significantly in format, including their correspondence with DSM criteria for PTSD, the dimension of symptom severity they emphasize (e.g. subjective distress, functional impairment, frequency), and the time frame they assess (past week, past month). Finally, they vary in the quality of their psychometric properties. Any of these characteristics, alone or in combination with characteristics of different samples, could affect their correlation with the CAPS. In general, in PTSD research, as in other areas of psychopathology, the diagnostic standard is a clinical interview because interviewers can clarify as needed, ask for examples, observe clinically relevant behaviors, and evaluate potential response bias. Most importantly, with an interview it is ultimately the clinician who makes the final rating, not the participant.

This, then, raises a third question. What measure should serve as the criterion for evaluating the diagnostic utility of the CAPS? Part of the problem is that there isn't another single measure that has been widely accepted as a criterion measure of PTSD. The SCID PTSD module comes the closest, but there is evidence suggesting it may not be as reliable as the CAPS, which sets an upper limit on how well the CAPS can perform in predicting it. In fact, as Weathers et al. (1999b) have argued, the CAPS appears to be more strongly associated with the SCID PTSD module than the SCID PTSD module is with itself. Another possibility might be to use a multiple converging measures approach, such as was used in the National Vietnam Veterans Readjustment Study (NVVRS; Kulka et al., 1990), or the so-called LEAD standard approach proposed by Spitzer and colleagues. Both approaches could readily be applied to the CAPS and would provide valuable new information.

Treatment Outcome Studies

Design and Analysis Issues

In this section we describe pharmacological and psychosocial treatment outcome studies that employed the CAPS as a primary outcome measure. Our main focus in this section is on the ability of the CAPS to detect genuine changes in PTSD symptom severity in the context of a clinical intervention. A key question addressed in this section is this: What empirical results would constitute evidence supporting the claim that the CAPS is in fact sensitive to change? We hypothesize four results we would expect to occur in a treatment outcome study if this claim is true. First, we would expect to find a reduction in CAPS scores from pre-treatment to post-treatment. This should be true for virtually any intervention, for any of the following reasons:

1. Possible placebo effects.

2. Possible statistical regression (i.e., participants selected on basis of extreme scores tend to show less extreme scores on subsequent testing).

3. The fact that repeated assessment, particularly interview-based assessment, may be considered an intervention in and of itself since it includes many putative active ingredients of psychotherapy, including (a) a safe, professional interpersonal context; (b) therapeutic exposure and emotional and cognitive processing through disclosure of painful aspects of the trauma and trauma-related symptoms; and (c) education about PTSD symptoms and self-monitoring.

Second, if a study includes one or more comparison groups, there should be greater improvement in the group or groups that receive a more potent treatment or a treatment with more putative active ingredients of therapy. Third, changes on the CAPS should parallel changes in other measures of PTSD. Finally, if the active therapy ingredient targets PTSD specifically then the CAPS should show greater reduction relative to measures of other constructs such as depression, anxiety, and global distress and

impairment.

In reviewing these studies we focused only on data related specifically to the CAPS. It was not our intent to address the effectiveness of pharmacological or psychosocial treatments for PTSD per se, or to rigorously critique the research methodology of the various studies. Nonetheless, within this limited scope of our review, we identified several issues regarding the reporting of CAPS data that required several decisions about how to extract and summarize CAPS-related results and present them in a standard format. First, studies varied considerably in terms of the outcome measures they included and how the data were reported and analyzed, differing on: (a) which CAPS scores were included (e.g., frequency, intensity, or severity scores for individual items, for the three symptom clusters, or for the syndrome as a whole); (b) which additional measures were included; (c) how scores were presented (e.g., means, totals); (d) how change was quantified (e.g., change scores, percent change, statistical significance, effect size, graphic presentation only); and (e) how complete the data analyses were. In general, in response to this variability we tried to extract the results most relevant to the CAPS and present them as uniformly as possible. For the purposes of this review we used percent change as the primary metric for comparing results across studies and across instruments within the same study. This is a commonly reported metric, particularly in the pharmacology literature. It is easily calculated when not provided, readily comprehensible, and applicable for any type of study, from case studies to large randomized trials. Where possible, we identified or calculated percent change for the primary outcome variables in each of the studies. In addition, we included the results of statistical significance tests of key comparisons when they were provided.

Second, studies varied in terms of how many measurement points they included. All studies included assessments at pre-treatment and post-treatment, but others included assessments at screening, extended baseline, pre-treatment, post-treatment, additional intervals during treatment, and one or more long-term follow-ups. To simplify our presentation, whenever possible we examined only pre-post changes for all studies. These data were available for almost all studies and were sufficient as evidence of the sensitivity of the CAPS to clinical change. Also, in the studies that presented additional follow-up data, pre-post changes were generally sustained and sometimes continued to improve, so little would have been gained by examining additional assessment periods.

Third, there was some ambiguity regarding the terms investigators used to describe their study designs. Terms such as open trial, uncontrolled trial, and open label do not adequately characterize the essential aspects of the research designs they were used to describe, nor were they used consistently across studies. The questions we used as a guide in depicting the various research designs were:

1. Is the treatment condition known to the participant?
2. Is the treatment condition known to the assessor?
3. Is there at least one comparison condition?
4. Is the comparison condition within-subjects, as in a crossover design, or between-subjects, as in a randomized controlled trial?

Answers to these questions were not always stated explicitly, although the investigators may have intended to imply them by the labels they used to describe their studies. In particular, unless otherwise specified, we assumed that assessments were not blinded. Fourth, studies often did not explicitly identify which version of the CAPS was used. This could sometimes be inferred, but in general, unless there was some specific indication that the CAPS-2 / CAPS-SX was used, we assumed that the CAPS-1 / CAPS-DX was used. Finally, the final sample size often differed from the initial one due to attrition and inclusion/exclusion criteria. We report the sample size on which the final data analyses were based.

Pharmacological and Psychosocial Treatment Studies

In this section we review 10 pharmacological and 19 psychosocial treatment studies that used the CAPS as a primary outcome measure. These studies and their key findings relevant to the CAPS are presented in Tables 1 and 2. We consider the results with respect to the four issues outlined above

regarding evidence of sensitivity to clinical change, including within-groups effects (pre-post change), between-groups effects (differential change due to nature of intervention, e.g. drug versus placebo), change on the CAPS relative to change on other measures of PTSD, and change on the CAPS relative to measures of other constructs (e.g, anxiety, depression, global distress and functional impairment). Whenever possible we present the percent change values for each measure described in Tables 1 and 2. However, some studies only reported the results of significance tests and did not include actual values for one or more key measures. The studies in Tables 1 and 2 are arranged chronologically and numbered within each table. For ease of presentation in the following sections, we refer to studies by number rather than by author(s) and year.

Within-groups effects. Among the pharmacological studies there was a significant reduction in CAPS total score in eight of the nine studies that reported inferential statistics (Table 1, all but Study 5 reported significance levels; all of those but Study 8 were significant). Considering only participants who received a drug, for the nine studies that reported actual CAPS score values (all but Study 2) the reduction in CAPS total score ranged from 10-63%, with a median of 33%. The psychosocial studies yielded similar findings, with evidence of even greater improvement. There was a significant reduction in CAPS total score in 10 of the 13 studies that reported inferential statistics (Table 2, Studies 1-4, 6, 8, 10-13, 15, 16, 19 reported significance levels; all of those but Studies 1, 6, and 8 were significant). Considering the participants who received an active intervention and showed the most improvement, for the studies that reported actual CAPS score values (all but Studies 1, 8, and 16) the reduction in CAPS total score ranged from 19-100%, with a median of 50%.

Between-groups effects. Overall, there were relatively few controlled trials. Of the 10 pharmacological studies, only three were randomized, placebo-controlled trials (Table 1, Studies 2-4). Two of these (Studies 2 and 3) found significantly greater reduction in CAPS scores for the drug group relative to the placebo group. The third study (Study 4) found slightly greater improvement for drug versus placebo, although the effect was not significant. Similarly, of the 19 psychosocial studies, only six were randomized, controlled trials (Table 2, Studies 1, 4, 12, 13, 16, 19), although an additional three studies (two crossover designs and one program evaluation) included a comparison condition (Studies 8, 10, 17). Only two of the six randomized, controlled trials (Studies 12 and 16) found significant between-groups effect, with significantly greater reduction in CAPS scores for a more active, trauma-focused intervention than for a control condition. Two of the other four studies (Studies 4 and 13) found greater improvement for active interventions relative to control conditions, but the effects were not significant. Of the remaining two studies, Study 19 included two active interventions, which showed substantial, equivalent improvement, but no minimal intervention control condition; Study 1 employed a very brief intervention and found no within-groups or between-groups changes on any measures. Finally, in Study 10, a quasi-experimental program evaluation, between-groups differences were found among three types of PTSD inpatient programs.

CAPS versus other PTSD measures. In general, CAPS results matched the results for self-report PTSD measures, particularly the IES. Among the pharmacological studies, the CAPS had comparable results to the IES in three studies (Table 1, Studies 1, 4, and 6) and to the DTS in two other studies (Studies 5 and 9), with differences ranging from 0-7 percentage points. For the psychosocial studies differences between the CAPS and other PTSD measures were more variable and somewhat larger. Eight studies (Table 2, Studies 2, 4, 5, 11, 12, 15, 16, and 19) found a greater reduction on the CAPS relative to the IES, with differences ranging from 1-24 percentage points. On the other hand, four studies (Studies 3, 9, 14, and 18) found a greater reduction on the IES, with differences ranging from 7-26 percentage points. In addition, the CAPS showed a comparable or greater reduction relative to the PK scale (Study 1), the Mississippi Scale (Study 5), the Civilian Mississippi Scale (Study 15), the PSS (Study 9), the MPSS-SR

(Study 7), the PCL (Study 11), and the Penn Inventory (Study 19).

CAPS versus measures of depression. All but two of the pharmacological studies included a measure of depression, primarily the HAM-D and MADRS. Four studies (Table 1, Studies 1, and 8-10) found greater reduction on the HAM-D relative to the CAPS, with differences ranging from 2-13 percentage points. A fifth study (Study 2) found significant within-groups and between-groups effects for both the CAPS and the HAM-D, but did not report actual rating scale values. However, two studies (Studies 6 and 7) found greater reduction on the CAPS relative to the MADRS, with differences of 9 and 10 percentage points respectively. One study (Study 5) found slightly greater reduction on the CAPS relative to the BDI. The BDI was also included in 12 of 19 psychosocial studies, with 10 (Table 2, Studies 2, 6, 7, 11-13, 15, 16, 18, 19) finding greater reduction on the CAPS and two (Studies 5 and 9) finding equivalent reduction on the two scales. Except for a case study (Study 7), which found a 48% reduction on the CAPS and an 18% increase on the BDI, the greater reduction on the CAPS ranged from 2-18 percentage points.

CAPS versus measures of anxiety. Four pharmacological studies included the HAM-A. Three (Table 1, Studies 1, 8, and 10) found greater reduction on the HAM-A relative to the CAPS, with differences ranging from 2-9 percentage points. The fourth study found a five percentage point greater reduction on the CAPS. One pharmacological study (Study 5) found a greater reduction on the CAPS relative to the STAI-S. Three psychosocial studies (Table 2, Studies 5, 11, and 12) included the STAI –S, two included the STAI-T (Studies 11 and 12), and one (Study 19) included the BAI. In each case the CAPS showed greater reduction, ranging from 2-59 percentage points.

CAPS versus global measures of distress and impairment. In the pharmacological studies, reduction in CAPS scores were accompanied by global measures of functioning, including the CGI in the six studies that employed it (Table 1, Studies 3, 5, 7, 8-10), the CIS (Study 6), and a relatively stringent consensus definition of treatment response (Study 1). Five psychosocial studies (Table 2, Studies 2, 3, 7, 13, and 15) included the SCL-90 and one (Study 10) included the BSI. In each case the CAPS showed greater reduction, ranging from 3-28 percentage points. In contrast, two studies (Studies 3 and 19) found greater reduction on the GHQ relative to the CAPS, and one study (Study 10) found greater reduction on the ASI psychiatric score.

Discussion. The 29 treatment outcome studies reviewed in this section provide ample evidence of the sensitivity of the CAPS to clinical change. We summarize the results by returning to the four hypothesized results discussed at the outset of this section. First, there was clear and consistent evidence of within-groups effects in both the pharmacological and the psychosocial treatment studies. Stronger within-groups effects were found in the psychosocial studies. This could be due to the fact that with one exception the drugs used in the studies reviewed were all antidepressants, and their efficacy for treating PTSD has not been clearly established. The symptom relief they bring about may be due more to their antidepressant effects rather than to specific effects on PTSD symptoms such as reexperiencing and effortful avoidance. In contrast, all of the psychosocial interventions involved some type of trauma-specific component, and most included some form of direct therapeutic exposure or cognitive processing, which have been shown to have specific effects on PTSD symptoms. This finding could also be due to the fact that in general the psychosocial interventions involved considerably more patient-therapist contact than did the pharmacological trials.

Second, there was some evidence of between-groups effects, although relatively few studies included a comparison condition. Two of the three pharmacological trials with a placebo control found greater reduction on the CAPS in participants who received the drug. Results were more inconsistent for the psychosocial trials. Only two of the six randomized trials, plus one quasi-experimental program

evaluation, found a significant between-groups effect. However, two of the nonsignificant trials employed quite limited interventions, and a third trial compared two active interventions, exposure and cognitive restructuring. Clearly, more randomized, placebo-controlled trials are needed before this issue can be resolved.

Third, reduction in CAPS scores was mirrored by reduction in self-report measures of PTSD, particularly the IES. The CAPS showed a slightly greater reduction than the IES in 2 of 3 pharmacological studies and 7of 11 psychosocial studies, although the margins, especially in the pharmacological trials, were generally small. Fourth, there was some evidence of greater reduction on the CAPS than on measures of depression, anxiety, and global distress, particularly on self-report measures such as the BDI, STAI, and SCL-90.

Finally, a comment about the populations studied. Although the CAPS was developed in a male combat veteran population, and many of the early studies focused exclusively on this population, the CAPS has now been extended to increasingly diverse samples that include females and victims of various types of civilian trauma. Of the studies reviewed in this section, 11 of 29 included at least some females and 15 of 29 included at least some participants with civilian trauma.

Validity Evidence From Case-Control Designs

In this section we consider validity evidence from studies in which participants were designated as PTSD-positive ("cases") or PTSD-negative ("controls") based on the CAPS, then compared on some biological or psychological measure or experimental task. Such case-control studies were too numerous and diverse to summarize briefly. Instead, we describe several representative examples from different research domains to illustrate that groups formed on the basis of a CAPS diagnosis differ in conceptually meaningful ways on a variety of characteristics or behaviors.

The first example involves the psychophysiology of PTSD. Physiological reactivity to reminders of the trauma is a core symptom of PTSD, and a growing number of studies have found that individuals with PTSD show greater reactivity than those without PTSD in laboratory-based physiological assessments. Much of the early work was conducted with male combat veterans, but more recent studies have examined male and female victims of civilian trauma. Blanchard et al. (1996a) used the CAPS to classify 105 male and female motor vehicle accident victims as PTSD, subsyndromal PTSD, and non-PTSD. They also included a control group of 54 participants who had not experienced an accident. They found that compared to participants without PTSD, those with PTSD showed a significantly greater increase in heart rate in response to brief audiotapes depicting each participant's unique traumatic experience. They also found that an increase of two beats per minute had reasonable diagnostic utility, yielding 69% sensitivity and 78% specificity among accident victims.

The second example comes from a more recent line of research on auditory event-related potentials (ERP) in PTSD. Several different investigators have documented abnormal ERPs in individuals with PTSD, and have suggested that such characteristic responses may be associated with the attention and concentration difficulties often seen in PTSD. Metzger et al. (1997) used the CAPS to classify male Vietnam combat veterans as PTSD or non-PTSD groups, then further divided the PTSD participants into medicated and unmedicated groups. Administering a three-tone auditory "oddball" task, they found significantly smaller P3 amplitudes in the unmedicated PTSD group, relative to the medicated PTSD group and the non-PTSD controls.

The third example involves research on the association of chronic PTSD and physical health problems. Beckham et al. (1998) used the CAPS to classify 276 male Vietnam combat veterans as PTSD or non-PTSD, then assessed participants' current health status and reviewed their medical records. Health measures included health complaints, current and lifetime physical conditions, number of physician-rated medical categories, and total number of physician-rated illnesses. After controlling for a variety of potentially confounding third variables, including age, socioeconomic status, ethnicity, combat exposure, alcohol problems, and smoking history, they found that veterans with PTSD had significantly more health

problems across all indicators compared to veterans without PTSD.

The last example represents an effort to identify potential risk factors for PTSD. Yehuda et al. (1995) used the CAPS to classify a community sample of 72 Nazi concentration camp survivors as PTSD or non-PTSD. They also included a comparison group of 19 demographically matched participants who had not experienced the Holocaust. The purpose of the study was to examine the relationships among lifetime trauma history, recent stressful life events, and severity of current PTSD symptoms. As expected, Yehuda et al. found that Holocaust survivors with PTSD had greater lifetime trauma exposure and more recent stressful life events than did survivors without PTSD or comparison participants. Using the CAPS as a continuous measure of PTSD symptom severity, they found that lifetime trauma was significantly associated with avoidance and hyperarousal, but not with reexperiencing, within a combined sample of all Holocaust survivors. In a similar analysis, they found that recent stressful life events were significantly associated with all three CAPS symptom clusters.

These examples, and the other case-control studies we did not discuss, provide additional evidence that the CAPS is a valid measure of PTSD diagnostic status and symptom severity. They demonstrate that when the CAPS is used to classify trauma-exposed individuals as PTSD or non-PTSD, the resulting groups differ significantly in a theoretically consistent way on key dependent variables.

General Discussion and Recommendations

In the ten years since it was developed, the CAPS has proven to be a psychometrically sound, practical, and flexible structured interview that is well-suited for a wide range of clinical and research applications in the field of traumatic stress. Moreover, it has been successfully used with many different traumatized populations. It has excellent reliability, yielding consistent scores across items, raters, and testing occasions. There is also considerable validity evidence supporting the use of the CAPS as a measure of PTSD diagnostic status and symptom severity. Evidence of content validity derives first from its direct correspondence with the DSM-IV diagnostic criteria for PTSD, and second from the fact that it was developed by experts in the field of traumatic stress and revised based on feedback from many clinicians and investigators who used it in real-world settings. Evidence from a growing number of psychometric investigations indicates it has strong convergent and discriminant validity, strong diagnostic utility, and is sensitive to clinical change. In addition, factor analyses, especially confirmatory factory analyses, have shown that the factor structure of the CAPS corresponds well to current conceptualizations of PTSD. Finally, when the CAPS is used in case-control designs, individuals designated as PTSD differ from those without PTSD in predictable, theoretically meaningful ways. Clearly more research on the CAPS is needed, but at this point the CAPS is the most extensively investigated structured interview for PTSD.

Criticism of the CAPS tends to focus on three concerns. The first concern is that the CAPS is cumbersome and lengthy. In response, the CAPS clearly is longer on paper than other PTSD interviews, but it doesn't necessarily take longer to administer. Most of the CAPS questions are optional probes, only some of which would likely be administered during a given interview. A standard administration of the CAPS involves asking the initial probe under frequency for each item. With an articulate, motivated respondent this single question may elicit all the information necessary to rate both the frequency and intensity of a given symptom. All other probes are to be used only if: (a) a response is incomplete, vague, confusing, or in some way insufficient to make a rating, and therefore needs to be clarified; or (b) the respondent doesn't understand what is being asked.

In our experience, even with ideal respondents, some degree of clarification is inevitable. To enhance uniformity of administration, we have included a number of followup probes that address the most common points of clarification. This reduces variability due to idiosyncratic questioning across different interviewers and provides a helpful structure for less experienced interviewers. Further, the CAPS was designed as a comprehensive yet flexible instrument that would meet the demand of almost any PTSD assessment task, including diagnosis, evaluating symptom severity, and conducting a

functional analysis of symptoms for case conceptualization and treatment planning. Therefore, in some assessment contexts may opt not to assess Criterion A, elicit descriptive examples of symptoms, administer the global ratings or the guilt and dissociation items, or rate lifetime PTSD.

The second concern, closely related to the first, is that the CAPS is too complicated and difficult to learn. In response, our own experience, based on dozens of training sessions, is that after a two-hour orientation trainees naïve to the CAPS can make highly reliable ratings of a role-played interview. With some self-study and a few practice interviews, they can achieve a uniform, clinically sensitive administration. CAPS trainees, including those with little or no experience with structured interviews or assessing PTSD, typically find that the CAPS is very straightforward to learn. In fact, less experienced interviewers tend to have the most favorable responses because they appreciate the structure the CAPS provides.

The third concern centers on the question of whether frequency and intensity ratings overlap to such an extent as to be essentially redundant. Clearly, they appear to be strongly correlated at the syndrome level and even at the symptom cluster level. At the item level, however, the correlations between frequency and intensity are moderate, suggesting that they measure correlated but distinct dimensions. We have several responses to this concern. First, the separate assessment of frequency and intensity explicitly defines what is meant by symptom severity, thereby reducing variability in clinical judgment, especially among less experienced interviewers. Second, this is a meaningful, theoretical distinction, employed successfully for example in the substance abuse literature, where typologies of drinkers are based on how often a person drinks, as well as how much they consume at any given setting. Third, adding frequency and intensity together yields a 9-point scale (0-8) that allows finer gradations of severity. This increases variance attributable to individual differences, thereby avoiding a restriction of range that could lower estimates of reliability and validity. Fourth, it allows the assessment of the differential impact of treatment on the frequency versus the intensity of symptoms.

Last, we close with some recommendations for the use of the CAPS in clinical research and the presentation of CAPS data in empirical reports. First, for newly initiated research, investigators should use what is now the sole version of the CAPS, the combined DSM-IV version, and explicitly identify it as such. For research already underway or completed, investigators should explicitly identify the version used, either the CAPS-1 or CAPS-2 (DSM-III-R versions) or the CAPS-DX or CAPS-SX (DSM-IV versions). Also, if the CAPS is used as a diagnostic measure, investigators should specify the scoring rule used to obtain a diagnosis. Second, investigators should briefly specify the experience and training of CAPS interviewers, both in terms of their general background in psychopathology and structured interviewing, and in terms of their specific experience with the CAPS. Also, whenever possible they should attempt to collect and report reliability data on the interviewers and participants involved. Even something as modest as interrater reliability on a small number of audiotaped interviews is helpful for documenting the quality of the CAPS data.

Third, investigators should take greater advantage of the flexibility of the CAPS in analyzing their data. Some examples include: (a) using multiple CAPS scoring rules and comparing the results for lenient, moderate, and stringent rules; (b) using the CAPS as both a dichotomous and a continuous measure, reporting not only diagnostic status but symptom severity scores, which would be valuable for comparing findings across studies; (c) breaking out CAPS symptom severity scores into the three DSM-IV symptom clusters and examining the results by cluster; (d) examining the symptom clusters further by separating Cluster C into effortful avoidance (C1 and C2) and emotional numbing; and (e) dividing scores even further into frequency, intensity, and severity scores for each of the symptom clusters. Finally, although considerable progress has been made in the development and evaluation of PTSD assessment measures, including the CAPS, reliance on a single instrument should be avoided. We advocate multimodal assessment of PTSD, an approach that relies on converging evidence from multiple sources, and we encourage investigators to include multiple measures of PTSD and comorbid disorders whenever possible.

References

American Psychological Association. 1999. Standards for Educational and Psychological Testing. Washington DC: Author.

Baker DG, Diamond BI, Gillette GM, Hamner MB, Katzelnick D, Keller TW, Mellman TA, Pontius EB, Rosenthal M, Tucker P, Van der Kolk BA, Katz RJ. 1995. A double-blind, randomized, placebo-controlled, multi-center study of brofaromine in the treatment of post-traumatic stress disorder. Psychopharmacology 122:386-389.

Beck AT, Epstein N, Brown G, Sterr RA. 1988. An inventory for measuring clinical anxiety: Psychometric properties. J Consult Clin Psychol 56:893-897.

Beckham JC, Moore SD, Feldman ME, Hertzberg MA, Kirby AC, Fairbank JA. 1998. Health status, somatization, and severity of posttraumatic stress disorder in Vietnam combat veterans with posttraumatic stress disorder. Am J Psychiatry 155:1565-1569/

Blake DD, Weathers FW, Nagy LM, Kaloupek DG, Klauminzer G, Charney DS, Keane TM. 1990. A clinician rating scale for assessing current and lifetime PTSD: the CAPS-1. Behav Ther 13:187-188.

Blake DD, Weathers FW, Nagy LM, Kaloupek DG, Gusman FD, Charney DS, Keane TM. 1995. The development of a Clinician-Administered PTSD Scale. J Trauma Stress 8:75-90.

Blanchard EB, Hickling EJ, Taylor AE, Forneris CA, Loos WR, Jaccard J. 1995a. 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. Behav Res Ther 33:471-475.

Blanchard EB, Hickling EJ, Taylor AE, Loos WR. 1995b. Psychiatric morbidity associated with motor vehicle accidents. J Nerv Ment Dis 183:495-504.

Blanchard EB, Hickling EJ, Buckley TC, Taylor AE, Vollmer A, Loos WR. 1996a. Psychophysiology of posttraumatic stress disorder related to motor vehicle accidents: Replication and extension. J Consult Clin Psychol 64:742-751.

Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. 1996b. Psychometric properties of the PTSD Checklist (PCL). Behav Res Ther 34:669-673.

Boudewyns PA, Hyer, Leon A (Lee). 1996. Eye movement desensitization and reprocessing (EMDR) as treatment for post-traumatic stress disorder (PTSD). Clinical Psychology and Psychotherapy 3:185-195.
Boudewyns PA, Stwertka SA, Hyer, Leon A (Lee), Albrecht JW, Sperr EV. 1993. Eye movement desensitization for PTSD of combat: a treatment outcome pilot study. Behav Ther 16:29-33.

Bouwer C, Stein DJ. 1998. Survivors of torture presenting at an anxiety disorders clinic: symptomatology and pharmacotherapy. J Nerv Ment Dis 186:316-318.

Buckley TC, Blanchard EB, Hickling EJ. 1998. A confirmatory factor analysis of posttraumatic stress symptoms. Behav Res Ther 36:1091-1099.

Busuttil W, Turnbull GJ, Neal LA, Rollins JW, West AG, Blanch N, Herepath R. 1995. Incorporating psychological debriefing techniques within a brief group psychotherapy programme for the treatment of

post-traumatic stress disorder. Br J Psychiatry 167:495-502.

Cañive JM, Clark RD, Calais LA, Qualls CR, Tuason VB. 1998. Bupropion treatment in veterans with posttraumatic stress disorder: an open study. J Clin Psychopharmacol 18:379-383.

Carlson JG, Chemtob CM, Rusnak K, Hedlund NL. 1996. Eye movement desensitization and reprocessing treatment for combat PTSD. Psychotherapy 33:104-113.

Carlson JG, Chemtob CM, Rusnak K, Hedlund NL, Muraoka MY. 1998. Eye movement desensitization and reprocessing (EMDR) treatment for combat-related posttraumatic stress disorder. J Trauma Stress 11:3-24.

Clark RD, Cañive JM, Calais LA, Qualls CR, Tuason VB. 1999. Divalproex in posttraumatic stress disorder: an open-label clinical trial. J Trauma Stress 12:395-401.

Conlon L, Fahy TJ, Conroy RM. 1999. PTSD in ambulant RTA victims: a randomized controlled trial of debriefing. J Psychosom Res 46:37-44.

Davidson JRT, Book SW, Colket JT, Tupler LA, Roth SH, David D, Hertzberg MA, Mellman TA, Beckham JC, Smith RD, Davison RM, Katz RJ, Feldman ME. 1997. Assessment of a new self-rating scale for posttraumatic stress disorder. Psychol Med 27:153-160.

Derogatis LR. 1983. SCL-90-R administration, scoring, and procedures manual-II for the revised version. Towson, MD: Clinical Psychometric Research

Fleming MP, Difede J. 1999. Effects of varying scoring rules of the Clinician Administered PTSD Scale (CAPS) for the diagnosis of PTSD after acute burn injury. J Trauma Stress 12:535-542.

Fontana A, Rosenheck RA. 1997. Effectiveness and cost of the inpatient treatment of posttraumatic stress disorder: comparison of three models of treatment. Am J Psychiatry 154:758-765.

Frueh BC, Turner SM, Beidel DC, Mirabella RF, Jones WJ. 1996. Trauma Management Therapy: a preliminary evaluation of a multicomponent behavioral treatment for chronic combat-related PTSD. Behav Res Ther 34:533-543.

Hall CA, Henderson CM. 1996. Cognitive processing therapy for chronic PTSD from childhood sexual abuse: a case study. Counselling Psychology Quarterly 9:359-371.

Hertzberg MA, Feldman ME, Beckham JC, Davidson JRT. 1996. Trial of trazodone for posttraumatic stress disorder using a multiple baseline group design. J Clin Psychopharmacol 16:294-298.

Hamilton, M. 1960. A rating scale for depression. J Neurol Neurosurg Psychiatry 23:56-62.

Hamilton, M. 1969. Diagnosis and ratings of anxiety. Br J Psychiatry 3: 76-79.

Hertzberg MA, Feldman ME, Beckham JC, Moore SD, Davidson JRT. 1998. Open trial of nefazodone for combat-related posttraumatic stress disorder. J Clin Psychiatry 59:460-464.

Hickling EJ, Blanchard EB. 1997. The private practice psychologist and manual-based treatments: post-traumatic stress disorder secondary to motor vehicle accidents. Behav Res Ther 35:191-203.

Hovens JEJM, Van der Ploeg HM, Bramsen I, Klaarenbeek MTA, Schreuder BJN, Rivero VV. 1994. The development of the Self-Rating Inventory for Posttraumatic Stress Disorder. Acta Psychiatr Scand 90:172-183.

Hyer, Leon A (Lee), Summers MN, Boyd S, Litaker M, Boudewyns PA. 1996. Assessment of older combat veterans with the Clinician-Administered PTSD Scale. J Trauma Stress 9:587-593.

Katz RJ, Lott MH, Arbus P, Crocq L, Herlobsen P, Lingjaerde O, Lopez G, Loughrey, Gerry C (Gerard), MacFarlane DJ, McIvor R, Mehlum L, Nugent D, Turner SW, Weisæth L, Yule W. 1994-1995. Pharmacotherapy of post-traumatic stress disorder with a novel psychotropic. Anxiety 1:169-174.

Keane TM, Caddell JM, Taylor KL. 1988. Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: three studies in reliability and validity. J Consult Clin Psychol 56:85-90.

Keane TM, Fairbank JA, Caddell JM, Zimering RT, Taylor KL, Mora CA. 1989. Clinical evaluation of a measure to assess combat exposure. Psychol Assess 1:53-55.

Keane TM, Malloy PF, Fairbank JA. 1984. Empirical development of an MMPI subscale for the assessment of combat-related posttraumatic stress disorder. J Consult Clin Psychol 52:888-891.

King DW, Leskin GA, King LA, Weathers FW. 1998. Confirmatory factor analysis of the Clinician-Administered PTSD Scale: evidence for the dimensionality of posttraumatic stress disorder. Psychol Assess 10:90-96.

Kulka RA, Schlenger WE, Fairbank JA, Hough RL, Jordan BK, Marmar CR, Weiss DS. 1990. The National Vietnam Veterans Readustment Study: tables of findings and technical appendices. New York: Brunner/Mazel.

Lazrove S, Triffleman EG, Kite L, McGlashan TH, Rounsaville B. 1998. An open trial of EMDR as treatment for chronic PTSD. Am J Orthopsychiatry 68:601-608.

Lubin H, Loris M, Burt J, Johnson DR. 1998. Efficacy of psychoeducational group therapy in reducing symptoms of posttraumatic stress disorder among multiply traumatized women. Am J Psychiatry 155:1172-1177.

Marks IM, Lovell K, Noshirvani H, Livanou M, Thrasher S. 1998. Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring: a controlled study. Arch Gen Psychiatry 55:317-325.

Marmar, CR, Weiss DS, Metzler TJ. 1997. The Peritraumatic Dissociative Experiences Questionnaire. In: Wilson JP, Keane TM, editors. Assessing psychological trauma and PTSD. New York: Guilford Press p 412-428.

Metzger LJ, Orr SP, Lasko NB, Pitman RK. 1997. Auditory event-related potential to tone stimuli in combat-related posttraumatic stress disorder. Biol Psychiatry 42:1006-1015.

Nagy LM, Morgan CA, Southwick SM, Charney DS. 1993. Open prospective trial of fluoxetine for posttraumatic stress disorder. J Clin Psychopharmacol 13:107-113.

Nagy LM, Blake DD, Schnurr P, Southwick SM, Charney D, Weathers F, Horner B. 1999. The Clinician-Administered PTSD Scale – Weekly Version (CAPS-2): Reliability and validity. Manuscript

submitted.

Neal LA, Busuttil W, Herepath R, Strike PW. 1994. Development and validation of the computerized Clinician Administered Post-Traumatic Stress Disorder Scale-1-Revised. Psychol Med 24:701-706.

Neal LA, Hill N, Hughes JC, Middleton A, Busuttil W. 1995. Convergent validity of measures of PTSD in an elderly population of former prisoners of war. Int J Geriatric Psychiatry 10:617-622.

Neal LA, Shapland W, Fox C. 1997. An open trial of moclobemide in the treatment of post-traumatic stress disorder. Int Clin Psychopharmacol 12:231-237.

Orr SP. 1997. Psychophysiologic reactivity to trauma-related imagery in PTSD: diagnostic and theoretical implications of recent findings. Ann Ny Acad Sci 821:114-124.

Pantalon MV, Motta RW. 1998. Effectiveness of anxiety management training in the treatment of posttraumatic stress disorder: a preliminary report. J Behav Ther Exp Psychiatry 29:21-29.

Pitman RK, Orr SP, Altman B, Longpre RE, Poiré RE, Macklin ML. 1996. Emotional processing during eye movement desensitization and reprocessing therapy of Vietnam veterans with chronic posttraumatic stress disorder. Compr Psychiatry 37:419-429.

Radnitz CL, Schlein IS, Walczak S, Broderick CP, Binks TM, Tirch DD, Willard J, Perez-Strumolo L, Festa J, Lillian LB, Bockian N, Cytryn A, Green L. 1995. The prevalence of posttraumatic stress disorder in veterans with spinal cord injury. SCI Psychosocial Process 8:145-149.
Rothbaum BO, Hodges L, Alarcón RD, Ready DJ, Shahar F, Graap K, Pair J, Hebert P, Gotz D, Wills B, Baltzell D. 1999. Virtual reality exposure therapy for PTSD Vietnam veterans: a case study. J Trauma Stress 12:263-271.

Shalev AY, Freedman SA, Peri T, Brandes D, Sahar T. 1997. Predicting PTSD in trauma survivors: prospective evaluation of self-report and clinician-administered instruments. Br J Psychiatry 170:558-564.

Spielberger CD, Gorsuch RL, Lushene RE. 1970. STAI manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Standards for educational and psychological testing. 1999. Washington DC: American Educational Research Association. 194 p.

Tarrier N, Pilgrim H, Sommerfield C, Faragher B, Reynolds M, Graham E, Barrowclough C. 1999. A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress disorder. J Consult Clin Psychol 67:13-18.

Taylor S, Kuch K, Koch WJ, Crockett DJ, Passey G. 1998. The structure of posttraumatic stress symptoms. J Abnorm Psychol 107:154-160.

Thompson JA, Charlton PFC, Kerry R, Lee D, Turner SW. 1995. An open trial of exposure therapy based on deconditioning for post-traumatic stress disorder. Br J Clin Psychol 34:407-416.

Thrasher SM, Lovell K, Noshirvani M, Livanou M. 1996. Cognitive restructuring in the treatment of post-traumatic stress disorder: two single cases. Clin Psychol Psychotherapy 3:137-148.

Van der Kolk BA, Dreyfuss D, Michaels MJ, Shera D, Berkowitz R, Fisler RE, Saxe GN. 1994. Fluoxetine in posttraumatic stress disorder. J Clin Psychiatry 55:517-522.

Weathers FW, Litz BT, Herman DS, Huska JA, Keane TM. 1993. The PTSD Checklist (PCL): Reliability, Validity, and Diagnostic Utility. Paper presented at the 9th Annual Meeting of ISTSS.

Weathers FW, Ruscio AM, Keane TM. 1999. Psychometric properties of nine scoring rules for the Clinician-Administered Posttraumatic Stress Disorder Scale. Psychol Assess 11:124-133.

Yehuda R, Kahan B, Schmeidler J, Southwick SM, Wilson S, Giller EL. 1995. Impact of cumulative lifetime trauma and recent stress on current posttraumatic stress disorder symptoms in Holocaust survivors. Am J Psychiatry 152:1815-1818.

Zlotnick C, Davidson JRT, Shea MT, Pearlstein T. 1996. Validation of the Davidson Trauma Scale in a sample of survivors of childhood sexual abuse. J Nerv Ment Dis 184:255-257.

Table 1. Summary of CAPS findings from pharmacological treatment studies of posttraumatic stress disorder.

Authors (year)	Participants	Design	Drug	Duration	Key CAPS-related findings
1. Nagy, Morgan, Southwick, & Charney (1993)	Male combat veterans (N=19)	1. Non-blinded, uncontrolled 2. CAPS-2	Fluoxetine	10 weeks	1. Significant reduction in CAPS-2 total score (34%) 2. Comparable reduction on IES (39%) 3. Somewhat larger reduction on HAM-D (47%) and HAM-A (41%) 4. With response defined as 50% reduction in CAPS total score, a 2-point improvement on CAPS global severity rating, and consensus of two clinicians, 7 participants (37%) had good response, 5 (26%) had partial response, 7 (37%) did not respond
2. van der Kolk, Dreyfuss, Michaels, Shera, Berkowitz, Fisler, & Saxe (1994)	1. Civilians with mixed trauma (N=23, 12 male / 23 female) 2. Combat veterans and civilians with mixed trauma (N=24, 23 male / 1 female)	Double-blind, randomized, placebo-controlled	Fluoxetine	5 weeks	1. Significantly greater reduction in total CAPS score for drug relative to placebo, after adjusting for initial CAPS score and site 2. Greater reduction in CAPS total score for civilian sample relative to veteran sample 3. Significant reduction in numbing and hyperarousal symptoms but not reexperiencing or avoidance 4. Significantly greater reduction in HAM-D score for drug relative to placebo
3. Katz, Lott, Arbus, Crocq, Herlobsen, Lingjaerde, Lopez, Loughrey, MacFarlane, McIvor, Mehlum, Nugent, Turner, Weisaeth, & Yule (1994/1995)	Combat veterans and civilians with mixed trauma (N=45, 34 male / 11 female)	Double-blind, randomized, placebo-controlled, multi-center	Brofaromine	14 weeks	1. Significant reduction in CAPS total score for both groups (drug=48%, placebo=29%), with significant between-groups difference 2. 55% of drug group and 26% of placebo group no longer met diagnostic criteria for PTSD 3. On CGI, drug group had significantly greater mean improvement and more participants rated as very much

					improved
4. Baker, Diamond, Gillette, Hamner, Katzelnick, Keller, Mellman, Pontius, Rosenthal, Tucker, van der Kolk, & Katz (1995)	Combat veterans and civilians with mixed trauma (N=114, 92 male / 22 female)	Double-blind, randomized, placebo-controlled, multi-center	Brofaromine	10 weeks	1. Significant reduction in CAPS total score for both groups (drug=33%, placebo=31%), but no between-groups difference 2. Comparable results for IES, with somewhat smaller reduction in IES total score in both groups (26%) and no between-groups difference 3. No between-groups difference on DTS or Physician's Global Evaluation (within-groups analyses not presented)
5. Hertzberg, Feldman, Beckham, & Davidson (1996)	Male combat veterans (n=6)	Multiple baseline, open label but assessment blind	Trazodone	4 months	1. Reduction in CAPS total score (15%) 2. Comparable reduction on DTS (15%) 3. Somewhat smaller reduction on BDI (10%), little change on STAI-S (+1%) 4. 4 of 6 participants rated as much improved on CGI, 2 rated as minimally improved
6. Neal, Shapland, & Fox (1997)	Military personnel and civilians with mixed trauma (N=20, 18 male / 2 female)	1. Non-blinded, uncontrolled 2. Computerized CAPS, intensity scores only	Moclobemide	12 weeks	1. Significant reduction in computerized CAPS total score (50%) 2. Comparable reduction on IES (49%) 3. Somewhat smaller reduction on MADRS (41%), HAM-A (44%), and CIS (39%) 3. Computerized CAPS change score correlated .76 with IES change score, but only .31 with MADRS and .32 with HAM-A change scores
7. Bouwer & Stein (1998)	Male torture victims (N=14)	Routine clinical care, non-blinded, uncontrolled	Sertraline (n=9) Imipramine	8 weeks	1. Significant reduction in CAPS total score (63%) 2. Somewhat smaller reduction on MADRS (53%)

			(n=2) Fluoxetine (n=2) Clomipramine (n=1)		3. 12 of 14 participants rated as very much or much improved on CGI
8. Canive, Clark, Calais, Qualls, Tuason (1998)	Male combat veterans (N=14)	Routine clinical care, non-blinded, uncontrolled	Bupropion	6 weeks	1. Trend for reduction in CAPS total score (10%), significant reduction (16%) in CAPS hyperarousal score, but not in reexperiencing (+1%) or avoidance/numbing (9%) scores 2. 10 of 14 participants rated as very much or much improved on CGI 3. Significant reduction on HAM-D (26%) but not HAM-A (12%)
9. Hertzberg, Feldman, Beckham, Moore, & Davidson (1998)	Male combat veterans (N=10)	Non-blinded, uncontrolled	Nefazadone	12 weeks	1. Significant reduction in CAPS total score (32%) 2. Significant, somewhat smaller reduction on DTS (28%) 3. Significant reduction on HAM-D (34%) but not BDI (7%) 4. 10 of 10 participants rated as much improved or very much improved on CGI
10. Clark, Canive, Calais, Qualls, & Tuason (1999)	Male combat veterans (N=13)	Open label but assessment (except CGI) blind, uncontrolled	Divalproex	8 weeks	1. Significant reduction in CAPS total (18%), reexperiencing (21%), and hyperarousal (29%) scores, nonsignificant reduction in avoidance/numbing score (7%) 2. Significant, somewhat larger reduction on HAM-D (31%) and HAM-A (27%) 3. 11 of 13 participants rated as much improved or very much improved on CGI

Note: BDI=Beck Depression Inventory, CGI=Clinical Global Impressions, CIS=Clinician Impression of Severity, DTS=Davidson Trauma Scale, HAM-A=Hamilton Rating Scale for Anxiety, HAM-D=Hamilton Rating Scale for Depression, IES=Impact of Event Scale, MADRS=Montgomery-Asberg Depression Rating Scale, STAI=State-Trait Anxiety Inventory.

Table 2. Summary of CAPS findings from psychosocial treatment studies of posttraumatic stress disorder.

Authors (year)	Participants	Design	Intervention	# of sessions / duration	Key CAPS-related findings
1. Boudewyns, Stwertka, Hyer, Albrecht, & Sperr (1993)	Male combat veterans (N=20)	Randomized, controlled trial, assessments not blinded	1. EMD 2. exposure control 3. routine clinical care (group therapy without exposure)	2 90-minute EMD or exposure sessions in 2 weeks	1. No significant reduction in any CAPS symptom or symptom cluster scores 2. No significant reduction on Mississippi Scale or IES 3. No significant reduction in psychophysiological responding
2. Busuttil, Turnbull, Neal, Rollins, West, Blanch, & Herepath (1995)	Military personnel, veterans and civilians with mixed trauma (N=34, 28 male / 6 female)	Uncontrolled, assessments not blinded	Inpatient group therapy	12 days	1. Significant reduction in CAPS total intensity (54%), global improvement (59%), and global severity (55%) scores 2. Significant, somewhat smaller reduction on IES (42%) and PK (48%) 3. Significant, somewhat smaller reduction on SCL-90 (43%), and BDI (39%) 4. 26 of 34 (76%) participants no longer met PTSD diagnostic criteria
3. Thompson, Charlton, Kerry, Lee, & Turner (1995)	Civilians with mixed trauma (N=23, 17 male/ 6 female)	Uncontrolled, assessments not blinded	Multicomponent cognitive-behavioral protocol (imaginal and in vivo exposure, cognitive restructuring)	8 weekly sessions	1. Significant reduction in CAPS total score (35%) 2. Significant, somewhat larger reduction on IES (42%) 3. Comparable reduction on SCL-90 (38%), larger reduction on GHQ (61%)
4. Boudewyns & Hyer (1996)	Male combat veterans(N=61)	Randomized, controlled trial,	1. EMDR 2. exposure	5-7 EMDR or exposure	1. Significant reduction in CAPS total score for all three groups (EMDR=33%, exposure=21%, routine care=17%),

		assessments blinded	control 3. routine clinical care (group therapy without exposure)	sessions in 6 weeks	but no significant between-groups differences 2. No significant reduction on IES 3. Significant between-groups differences on POMS anxiety scale and heart rate reactivity, with EMDR and exposure group showing reduction in scores and no-exposure control group showing slight increase
5. Carlson, Chemtob, Rusnak, & Hedlund (1996)	Male combat veterans (N=4)	Single-subject replication series	EMDR	12 sessions, 2 sessions per week	1. At 3-month followup, reduction in CAPS total score across 4 participants ranged from 34-100%, with 3 of 4 showing > 80% improvement 2. Comparable reduction on IES (34-88%), but smaller reduction on Mississippi Scale (6-46%) 3. More variable outcome on BDI and STAI-S and STAI-T (1 participant showing slight increase on these scales, other 3 showing reduction of 50-100% reduction on BDI and 8-41% on STAI)
6. Frueh, Turner, Beidel, Mirabella, & Jones (1996)	Male combat veterans (N=11)	Uncontrolled, assessments not blinded	Multicomponent cognitive-behavioral protocol (education, imaginal and in vivo exposure, social skills training, anger management)	29 sessions in 17 weeks	1. Trend for reduction in CAPS total score (21%) 2. Significant reduction on HAM-A (31%), CGI (34%) and heart rate reactivity (14%) 3. No significant reduction on BDI, SPAI, or STAXI
7. Hall & Henderson (1996)	Female sexual abuse victim (N=1)	1. Case study 2. CAPS-2	Cognitive processing therapy	17 weekly sessions	1. Reduction in CAPS-2 total score (48%) 2. Smaller reduction on MPSS-SR (31%) 3. Somewhat smaller reduction on SCL-90 (26%), and

					slight increase on BDI (+18%)
8. Pitman, Orr, Altman, Longpre, Poire, & Macklin (1996)	Male combat veterans (N=17)	Crossover, assessments blinded	EMDR, with and without eye movements	12 weekly sessions (6 in each condition)	1. Little change in CAPS total score, with slight increase after eye movement condition and slight decrease after no eye movement condition 2. Comparable result for Mississippi Scale, with slight increase after both conditions, and mixed results for IES, with significant reductions for intrusion or avoidance subscale depending on condition and trauma memory evaluated 3. Significant reduction on SCL-90 in eye movement condition 4. Therapy integrity ratings significantly correlated with CAPS change score in both conditions (.55, .62), but with SCL-90 in eye movement condition only (.69)
9. Thrasher, Lovell, Noshirvani, & Livanou (1996)	Male physical assault victims (N=2)	Single-subject replication series	Cognitive restructuring	10 sessions	1. Substantial reduction in CAPS total score for both participants (67-90%) 2. Comparable reduction on IES (76-91%) and PSS (79-80%) 3. Comparable reduction on BDI (65-92%)
10. Fontana & Rosenheck (1997)	Male combat veterans (N=785)	Quasi-experimental program evaluation	1. Long-stay PTSD program 2. Short-stay PTSD program 3. General psychiatric unit	Variable (approximately 1-3 months)	1. Significant reduction in CAPS total score for all three programs (long-stay=13%, short-stay=19%, psychiatric=16%) 2. Significant between-groups effect, with veterans in short-stay PTSD programs and general psychiatric inpatient units showing greater improvement 3. No significant reduction on Mississippi Scale (long-stay=0%, short-stay=3%, psychiatric=3%)

					4. Significant, larger reduction on ASI psychiatric score (long-stay=24%, short-stay=26%, psychiatric=26%) and significant, smaller reduction on BSI (long-stay=2%, short-stay=11%, psychiatric=12%), both with significant between-groups effects similar to those for the CAPS
11. Hickling & Blanchard (1997)	Motor vehicle accident victims (N=10, 1 male / 9 female)	Uncontrolled trial, nonblinded assessments	Multi-component cognitive-behavioral protocol (education, relaxation, exposure, cognitive restructuring)	10 weekly sessions	1. Significant reduction in CAPS total score (68%) 2. Comparable reduction on IES (66%), significant but smaller reduction on PCL (39%) 3. Significant, somewhat smaller reduction on BDI (50%) and significant, smaller reduction on STAI-S (19%) and STAI-T (20%) 4. 5 of 8 participants with full PTSD and 1 of 2 with subsyndromal PTSD at pre-test no longer met diagnosis at post-test; 3 of 8 with full PTSD at pre-test were subsyndromal at post-test
12. Carlson, Chemtob, Rusnak, Hedlund, & Muraoka (1998)	Male combat veterans (N=35)	Randomized, controlled trial, non-blinded assessments except at 9-month followup	1. EMDR 2. Biofeedback-assisted relaxation 3. Routine clinical care	12 sessions in 6 weeks	1. Significant Group x Time interaction at 3-month followup, with EMDR group showing significantly greater reduction on CAPS total score (69%) compared to relaxation group (20%) 2. Similar pattern with smaller reduction on IES (EMDR= 45%, relaxation=14%) 3. Similar pattern with smaller reduction on BDI (EMDR=57%, relaxation= 22%), and substantially smaller reduction in both groups on STAI-S (EMDR=14% , relaxation=18%) and STAI-T (EMDR=22%, relaxation=11%) 4. Of participants completing first followup, 7 of 9 (78%) in EMDR group versus 2 of 9 (22%) in relaxation group no longer met PTSD diagnostic criteria

13. Conlon, Fahy, & Conroy (1998)	Motor vehicle accident victims , 1 week post-accident (N=40, 19 male / 21 female)	Randomized, controlled trial, non-blinded assessments	1. Debriefing 2. Monitoring (assessment-only control)	Single 30-minute debriefing session	1. Significant reduction in CAPS total score for total sample (53%; debriefing=70%, monitoring=36%), but no significant between-groups difference at follow-up 2. Interpretation of CAPS change scored is somewhat ambiguous because CAPS-2 used at baseline and CAPS-1 used at followup 3. Comparable reduction on IES (total sample=50%; debriefing=55%, monitoring=44%)
14. Lazrove, Triffleman, Kite, McGlashan, & Rounsaville (1998)	Civilians with mixed trauma (N=8, 2 male/ 6 female)	Uncontrolled trial, assessments conducted by non-treating research assistant	EMDR	3 weekly sessions	1. Substantial reduction in CAPS total score (70%) 2. Larger reduction on IES-R (87-.96% for intrusion, avoidance, hyperarousal subscales) 3. Comparable reduction on BDI (68%), smaller reduction on SCL-90 (42%) 4. All of the participants who completed treatment no longer met diagnostic criteria for PTSD
15. Lubin, Loris, Burt, & Johnson (1998)	Female victims of mixed civilian trauma	Uncontrolled trial, assessments conducted by non-treating research assistants	Trauma-focused, cognitive behavioral group therapy	16 weekly sessions	1. Significant reduction in CAPS total score (39%) 2. Significant, smaller reduction on Civilian Mississippi Scale (9%) and IES (16%) 3. Significant, somewhat smaller reduction on BDI (33%) and smaller reduction on DES (21%) and SCL-90 (23%)
16. Marks, Lovell, Noshirvani, Livanou, & Thrasher (1998)	Civilians with mixed trauma (N=87, 56 male/ 31 female)	1. Randomized, controlled trial, blinded assessments 2. CAPS-2	1. Imaginal and in vivo exposure 2. Cognitive restructuring 3. Exposure plus cognitive	10 sessions in an average of 16 weeks	1. Significant reduction in CAPS-2 total score, with effect sizes ranging from 1.30 to 2.00 for three active intervention groups and .60 for relaxation group 2. Significant between-groups effect for CAPS-2 total score, with greater reduction for three active intervention groups pooled versus relaxation group

			restructuing 4. Relaxation (placebo control)		3. Similar within-groups and between-groups results for IES (within-groups effect sizes from 1.30 to 1.50 for active intervention groups, .08 for relaxation group) 4. Similar within-groups and between-groups results for BDI (within-groups effects sizes from 1.20 to 1.70 for active intervention groups, .07 for relaxation group) 5. With improvement defined as > 2 SDs, 47-53% of participants in active intervention groups showed improvement on CAPS-2 total score, versus 15% in relaxation groups. Somewhat higher rates found for IES (50-60% for active intervention groups, 20% for relaxation group) 6. 63-75% of participants in active intervention groups versus 55% in relaxation group no longer met diagnostic criteria for PTSD
17. Pantalon & Motta (1998)	Male combat veterans (N=6)	1.Crossover, single-subject replication series, non-blinded assessments 2. CAPS-2	1. Implosive therapy (imaginal exposure) 2. Anxiety management training	12 weekly sessions	1. Reduction in CAPS-2 score (reexperiencing and avoidance only; hyperarousal scores not reported) ranged from 46-88% (M=71%) across the six participants 2. Lower but substantial reduction on PCL (8-100%, M=50% across the six participants; reexperiencing and avoidance only; hyperarousal scores not reported)
18. Rothbaum, Hodges, Alarcon, Ready, Shahar, Graap, Pair, Hebert, Gotz, wills, & Baltzell (1999)	Male combat veteran (N=1)	Case study	Virtual reality exposure	14 sessions, 2 sessions per week	1. Reduction in CAPS total score (34%) 2. Larger reduction on IES (45%) 3. Smaller reduction on BDI (19%) and STAXI-T (21%), substantially larger reduction on STAXI-S (63%)
19. Tarrier, Pilgrim, Sommerfield, Faragher,	Civilians with mixed trauma (N=62, 36 male/	Randomized, controlled trial, assessments	1. Imaginal exposure 2. Cognitive	Average of 10-12 sessions	1. Significant within-groups reduction in CAPS total score for both groups (32% for exposure group, 35% for cognitive therapy group), but no between-groups

Reynolds, Graham, & Barrowclough (1999)	26 female)	blinded	therapy	over 6 months	difference 2. Comparable within-groups reduction on IES (31-33% for intrusion, 25-34% for avoidance), somewhat smaller reduction on Penn Inventory (22-27%); no between-groups difference on either 3. Somewhat smaller within-groups reduction on BDI (27-31%) and BAI (23-25%); no between-groups difference 4. Comparable to somewhat larger within-groups reduction on GHQ (30-46%), but no between-groups difference 5. 59% of exposure group versus 42% of cognitive therapy group no longer met diagnostic criteria for PTSD

Note: ASI=Addiction Severity Index, BDI=Beck Depression Inventory, BSI=Brief Symptom Inventory, DES=Dissociative Experiences Scale, EMDR=Eye Movement Desensitization and Reprocessing, GHQ=General Health Questionnaire, HAM-A=Hamilton Rating Scale for Anxiety, IES=Impact of Event Scale, MPSS-SR=Modified PTSD Symptom Scale – Self-Report, PCL=PTSD Checklist, PK=Keane MMPI PTSD scale, POMS=Profile of Mood States, PSS=PTSD Symptom ScaleSCL-90=Symptom Checklist-90, SPAI=Social Phobia and Anxiety Inventory, STAI=State-Trait Anxiety Inventory, STAXI=State-Trait Anger Expression Inventory

Appendix #2

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

Adapted from: Weathers, F.W., Ruscio, A.M., & Keane, T.M. (1999). Psychometric Properties of the Nine Scoring Rules for the Clinician Administered Posttraumatic Stress Disorder Scale, *Psychological Assessment, 11,* 124-133.

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

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 roles 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-111-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

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 ClinicianAdministered 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.' 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.aubum.edu.

'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-111-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.' 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-111-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

	Sample				
	1	2	3	4	5
Variable	(n = 123)	(n = 24)	(n = 53)	(n = 67)	(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
BASS 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

Note. GED = Graduate Equivalency Diploma.

a Percentages summing to over 100%a 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 caters was used. Two of the three caters 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-111-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-111-R version of the CAPS. All other participants were administered the DSM-lV 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-111-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-111-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 DSMIII-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 90item 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 (Derogates, 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 DSMIII-R or DSM-IV algorithm for PTSD (one reexperiencing symptom, three avoidance and numbing symptoms, and two hyperarousal symptoms).

Frequency > 1/Intensity ? 2 (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 (SXCAG). 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/I12/TSEV65). This rule combines the F1112 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

	Sample	
Scoring rule	1 (n = 60)	2 (n = 24)
Rationally derived rules		
Frequency > 1/Intensity > 2’	.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’	.89	1.00
Total severity > 45	.85	.78
Total severity > 65a	.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-111-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-111-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 DST 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), K(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, K(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, K(.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 K(1) and K(0), K(.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 K(.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 K(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 r<(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	K(0)	K(.5)	K(1)
Rationally derived rules									
Frequency > 1/Intensity > 2	.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 > 65a	.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; K(0) = kappa coefficient representing quality of specificity; K(.5) = kappa coefficient representing quality of efficiency; K(1) = kappa coefficient representing quality of sensitivity.

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 FI/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 F1/12 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

	Sample	
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. ‘ 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 F1/12	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

Combined Research Clinical
Scoring rule (n=243) (n=571)

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., Chamey, 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 ClinicianAdministered 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, 1., 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: NeoKraepelinian 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-111* 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-11 (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,* 914923.

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. Psy*chological 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 A 1

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.

of

Table A2

Pairs for Dichotomizing Clinician-

is

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

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

counted as a symptom toward a PTSD diagnosis.

Received July 27, 1998

Revision received January 12, 1999
Accepted January 14, 1999