POSTTRAUMATIC STRESS DISORDER

A COMPREHENSIVE TEXT

EDITED BY

PHILIP A. SAIGH
Graduate School and University Center
of the City University of New York

J. DOUGLAS BREMNER
Yale University School of Medicine
and Veterans Administration
Connecticut Health Care System
West Haven Campus

ALLYN AND BACON
Boston • London • Toronto • Sydney • Tokyo • Singapore

CHAPTER 11

PSYCHOLOGICAL ASSESSMENT OF TRAUMATIZED ADULTS

FRANK W. WEATHERS, PH.D.
Auburn University

TERENCE M. KEANE, PH.D.
Boston Department of Veterans Affairs Medical Center
and Boston University School of Medicine

Over the last ten years, the development and psychometric evaluation of assessment tools for posttraumatic stress disorder (PTSD) has been one of the most productive areas of investigation in the field of traumatic stress. The remarkable progress that has been achieved can be illustrated by considering that in the mid-1980s, when investigators for the National Vietnam Veterans Readjustment Study (NVVRS) undertook their landmark epidemiological study of PTSD in Vietnam veterans, they were unable to identify a single measure that had been shown to be valid for distinguishing PTSD cases from non-cases (Kulka et al., 1991). Consequently, they conducted a preliminary validation study to identify appropriate measures of PTSD for use in the main study. Today, however, due to the pioneering work in the NVVRS and subsequent efforts by numerous other trauma researchers, there is an abundance of instruments to assess PTSD. More than two dozen well-validated measures are currently available, including structured interviews, questionnaires, and psychophysiological protocols, several of which, such as the Impact of Event Scale (IES; Horowitz, Wilner, & Alvarez, 1979) and the Mississippi Scale for Combat-Related PTSD (Mississippi Scale; Keane, Caddell, & Taylor, 1988), can be considered psychometrically mature on the basis of the extensive empirical literature they have inspired.

This rapid increase in the number of PTSD assessment tools has clearly had a positive impact on the field of traumatic stress. The availability of standardized measures has greatly facilitated the rapidly accumulating empirical literature regarding the clinical features, etiology, course, and treatment of PTSD, and has provided clinicians the means for evaluating the presence and severity of posttraumatic symptoms in their patients. Despite the gains that have been made, however, several problems remain. First, as we discuss later in the chapter, the assessment of trauma and PTSD raises many complex conceptual and methodological issues that have yet to be resolved or in some cases even fully articulated. Second, the assessment of PTSD has not yet become routine outside of trauma research centers, and continued outreach and education is needed to encourage more investigators and clinicians to add measures of trauma and PTSD to their assessment batteries.

Third, and somewhat paradoxically given the second point, there is a growing concern that the proliferation of PTSD assessment instruments is leading to fragmentation in the field. Instead of too few measures, as was the case a decade ago, today there simply may be too many. Despite the already large number of existing instruments, articles describing new measures of trauma, PTSD, or closely related constructs continue to appear regularly in the literature. While a few measures have been used successfully across different labs and trauma populations, most have not been widely adopted outside the original setting in which they were developed, and there is

no clear consensus about which measures should be included in a core PTSD assessment battery. The unfortunate result is that disparities in instrumentation across studies are more the rule than the exception in PTSD research. This hinders direct comparisons of findings across studies and undermines efforts to identify commonalities among survivors of different types of traumatic events.

To document the extent of this problem, we searched the PILOTS database (Lerner, 1994), examining entries for all empirical articles published in 1996 on PTSD in adults. This informal review led to several conclusions, which obviously are bounded by the information available in the PILOTS citations:

1. Encouragingly, the use of standard PTSD assessment instruments has become the norm, though articles based on non-standardized assessment procedures (e.g., unstructured interviews, study-specific measures, chart reviews) continue to be published. Of 189 articles that fit our search criteria, 157 (83 percent) used at least one standardized measure of PTSD.

2. There is considerable variability in the measures used to assess PTSD. We identified 31 different measures in our review, and this does not take into account multiple versions of established measures, measures that existed before 1996 but were not used in the studies we reviewed, measures that are too new to have appeared in the literature, or study-specific measures. The IES and the Mississippi Scale were the most commonly used measures overall, with 68 (36 percent) and 49 (26 percent) citations respectively, and the Clinician-Administered PTSD Scale (CAPS; Blake et al., 1990; Blake et al., 1995) was the most commonly used diagnostic interview, with 37 (20 percent) citations. Overall these findings reveal some degree of overlap in PTSD assessment batteries, but indicate that the field is far from consensus.

3. Too many investigators rely exclusively on self-report measures. Only 90 (48 percent) studies used a structured interview in their assessment battery, and this may be an overestimate since it was not always possible to determine if the PTSD module of a comprehensive interview such as the Structured Clinical Interview for DSM-III-R (SCID; Spitzer, Williams, Gibbon, & First, 1990) was administered. Although there are circumstances in which reliance on self-report measures may be the only feasible approach, a structured interview, preferably by an experienced clinician, is the foundation of any thorough assessment of PTSD.

4. Too few investigators use multiple measures, a practice that has often been advocated in the assessment of PTSD (e.g., Keane, Wolfe, & Taylor, 1987; Kulka et al., 1991). Only 75 studies (40 percent) used more than one standard measure, and only 29 (15 percent) used three or more.

5. Ironically, there is more uniformity in the assessment of comorbid problems, with the SCID, the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Luschene, 1970), and the Symptom Checklist-90 Revised (SCL-90-R; Derogatis, 1983) among the most commonly used instruments.

The shortcomings in the assessment of PTSD identified in our brief review are attributable to several factors. In some studies the data were collected prior to the widespread availability of well-validated measures, as in the case of secondary analyses of archival data sets. In other studies the investigators simply may not have been aware of high-quality measures that were available. In still other studies PTSD appeared to be peripheral to the primary research question, so little attention was given to assessing it.

All of these circumstances are understandable in a relatively new field and are easily addressed as the field progresses and new investigations are undertaken. What is of greater concern is that some investigators, perhaps feeling compelled by novel aspects of their research question or target population, continue to develop new PTSD instruments or to alter existing ones to create an assessment battery that will satisfy the unique demands of their study. The advantage of this decision is that the investigators have a measure that, at least in a face-valid way, is tailored to their specific assessment question and population. The disadvantages, however, are that the psychometric properties of the new measure are unknown and that the comparability with other relevant studies is compromised due to a lack of a common reference point. A

better approach, if a study-specific measure were considered essential, would be to include one or more standard measures in the assessment battery in order to maintain some direct link with other research. Given the variety of well-validated PTSD measures currently available, however, the use of non-standard measures is increasingly difficult to justify.

To address these and other concerns, a group of 45 leading trauma researchers, funded in a collaborative effort by the National Center for PTSD and the National Institutes for Mental Health, gathered in the fall of 1995 for a consensus conference on the assessment of PTSD. This meeting was similar in scope and intent to a consensus conference on panic, in which leading panic researchers defined key terms and identified acceptable instruments for assessing panic disorder (Shear & Maser, 1994). Five major areas were considered, including adult PTSD, childhood PTSD, traumatic stressors, comorbidity, and functional status. The intent of the conference was not to advance a set of rigid dictates, but rather to develop flexible, pragmatic guidelines that would promote the standardization of assessment in the field of traumatic stress, thereby enhancing comparability of research findings across different trauma populations and different settings. The conclusions from this conference, which will be summarized in a forthcoming manuscript and which we cite throughout the chapter, are an important first step toward much-needed uniformity in the assessment of trauma and PTSD.

In the remaining sections of this chapter we draw on the existing literature, on our own experiences developing and evaluating PTSD measures, and on recommendations from the consensus conference to describe the current status of PTSD assessment and challenges for the future. First, we discuss some of the central conceptual and methodological issues in the assessment of PTSD. Second, we outline the psychometric considerations involved in evaluating psychological tests. Third, we provide an overview of some of the most widely used PTSD interviews and self-report measures. Finally, we briefly describe different assessment scenarios that we have encountered in our work, outlining the assessment batteries used and the diagnostic decision-making involved in answering the assessment questions. This last section is intended as a cookbook which other investigators and clinicians can use to inform their own assessment practices. To the extent that a scenario fits a reader's specific assessment needs, the guidelines we provide can be adopted directly. But even in the case of a novel assessment question that does not correspond well to any of the scenarios we present, the examples in this section may provide a useful starting point toward a satisfactory solution.

No single chapter could summarize the vast literature on the assessment of PTSD; accordingly, this chapter has several limitations. First, we focus specifically on the assessment of PTSD and not on other stress-related syndromes such as acute stress disorder or complex PTSD (e.g., Herman, 1992). Second, we focus on PTSD symptoms, touching only briefly on the assessment of traumatic life events. Third, we focus on adults, since the assessment of PTSD in children is covered elsewhere in this volume (see chapter 10). Fourth, we focus on PTSD interviews and self-report measures. Information on other assessment approaches such as psychophysiological and other biological measures, neuropsychological tests, and projective tests can be found in a comprehensive new volume devoted to the assessment of PTSD (Wilson & Keane, 1997).

Finally, we focus on the diagnostic function of the assessment instruments we describe. It is important to note, however, that diagnosis is just one part of assessment, albeit a central part. As we have described elsewhere (e.g., Keane et al., 1987; Litz & Weathers, 1994), the assessment of trauma and PTSD is a complex process that unfolds in a rich interpersonal context. When properly conducted it can be both educational and therapeutic for the traumatized individual being assessed. The goals for different types of assessment vary, but in most clinical applications they would include looking beyond diagnosis to understand the impact of trauma on an individual's life, identifying and prioritizing specific targets for change, and offering clear feedback to the client.

CONCEPTUAL AND METHODOLOGICAL ISSUES IN THE ASSESSMENT OF PTSD

In this section we discuss some of the central issues regarding the assessment of PTSD in traumatized

adults. Many of these issues overlap, and distinctions among them are somewhat arbitrary, but we consider them separately to highlight the unique challenges that each poses. Though some involve more purely theoretical considerations, all have significant practical implications for the actual conduct of PTSD assessments in any clinical or research context. This is by no means an exhaustive list, but the issues described below are the ones we feel bear most directly on the development and utilization of standardized PTSD assessment tools.

Evolution of the PTSD Diagnostic Criteria

Since PTSD was first introduced as a formal diagnostic entity in the DSM-III in 1980, the diagnostic criteria have evolved considerably (see Saigh & Bremner, 1998; Wilson, 1995). Most of the changes occurred between the DSM-III and the DSM-III-R, including (a) expanding the number of symptoms from 12 to 17; (b) recasting symptoms into the three clusters of reexperiencing, avoidance and numbing, and hyperarousal; (c) revising and further explicating several criteria; and (d) dropping guilt as a criterion. More modest changes for the DSM-IV included (a) an extensive reworking of Criterion A (see next section); (b) moving cued physiological reactivity from the hyperarousal cluster to the reexperiencing cluster; (c) adding the requirement that symptoms must cause marked subjective distress or functional impairment; and (d) rewording and clarifying of some symptoms.

These changes in the diagnostic criteria reflect ongoing controversy and ambiguity regarding the core phenomenology of PTSD, which again is both legitimate and understandable in a newly emerging field. The problem is that changes have occurred so rapidly that there simply has not been enough time to fully explore one definition before the next is introduced. Combined with the proliferation of PTSD assessment instruments, the evolution of the PTSD criteria has hindered the accumulation of the data needed to address the most pressing questions empirically.

Fortunately, the diagnostic criteria for PTSD appear to have stabilized. Apart from the revised definition of what constitutes a traumatic event, there appears to be little substantive difference between the PTSD criteria in the DSM-III-R and the DSM-IV. To the extent that the new definition of a trauma is more explicit and therefore possibly more stringent, some individuals may not receive a PTSD diagnosis according to the DSM-IV because they fail to meet Criterion A. But given that an unequivocal traumatic event has occurred, the DSM-III-R and the DSM-IV yield virtually identical diagnostic results. A recent investigation from our lab suggests that only a very small fraction of individuals would be classified differently under DSM-III-R versus DSM-IV criteria (Weathers, Meron, & Keane, 1997).

From a practical assessment perspective, our view is that clinicians and investigators should always adopt the most current PTSD diagnostic criteria as soon as feasible. This means that new investigations should utilize measures that accurately reflect the DSM-IV criteria. Moreover, DSM-III-R diagnoses from investigations initiated prior to 1994 can be converted to DSM-IV diagnoses in most instances, allowing researchers to explore the implications of using different definitions of PTSD. The use of the DSM-III criteria, however, is now less than optimal and is only appropriate if the investigation involves a reanalysis of an archival data set, or if information regarding the DSM-IV criteria is also collected.

Despite the emerging consensus and stability in the PTSD diagnostic criteria, a remaining concern is whether these criteria are interpreted and utilized similarly by different clinicians and investigators. Although many criteria are clearly defined and relatively straightforward to assess, we and our colleagues have struggled to reach consensus regarding the parameters of more equivocal symptoms such as flashbacks, amnesia, and foreshortened future. Also, we are well aware that other clinicians and investigators wrestle with this issue, often developing idiosyncratic meanings for problematic criteria.

Ambiguity in the PTSD diagnostic criteria makes it difficult to inquire about symptoms and to clarify them for respondents, and contributes to unreliability in ratings of symptom severity. The text of the DSM-IV provides little explication of individual PTSD symptoms, and the resources that are available (e.g., Weiss, 1997) do not resolve all of the connotative issues. In developing the CAPS we tried to antic-

ipate typical points of confusion that occur during symptom inquiry and to provide standard follow-up prompts to address them. Nonetheless, whenever we provide consultation or training on PTSD assessment we are invariably struck by how much discussion and consensus building is needed regarding the definition of the diagnostic criteria, especially early in the implementation of a new assessment protocol.

The Criterion A Problem

One of the most vexing issues in the field of traumatic stress, and a focal point for much heated debate, is the so-called Criterion A problem, a broad label for several crucial questions regarding the proper role of traumatic events in the diagnosis of PTSD. Initially the Criterion A problem centered on questions such as how to define a traumatic event and whether experiencing a traumatic event should be a requirement for a PTSD diagnosis. These questions are still unresolved, but there appears to be a growing consensus that traumatic events can be distinguished from ordinary stressors (albeit with some inevitable imprecision and ambiguity) and that they serve a useful gatekeeping function that preserves the meaningfulness of the disorder as a distinct diagnostic entity and prevents trivialization of the suffering of survivors of overwhelming stressors. Concerns regarding Criterion A continue to focus on how to define a trauma, but have also shifted to other questions about how to measure traumatic events and whether to require a link between PTSD symptoms and a specific stressor (for more extensive discussion of the Criterion A problem, see Davidson & Foa, 1991; Kilpatrick & Resnick, 1993; Resnick, Kilpatrick, & Lipovsky, 1991; and Solomon & Canino, 1990).

Defining Traumatic Events. Some of the most marked changes in the PTSD diagnostic criteria have been in the definition of Criterion A, the requirement that an individual must have experienced a traumatic life event in order to qualify for a PTSD diagnosis. The definitions offered in the DSM-III and the DSM-III-R focused on the phrase "outside the range of usual human experience." As many observers have noted, however, this definition is unsatisfactory, in part because traumatic events are actually fairly common and in part because it misses the essential characteristic that all traumatic events have in common. As Herman (1992) has argued:

> *Traumatic events are extraordinary, not because they occur rarely, but rather because they overwhelm the ordinary human adaptations to life. Unlike commonplace misfortunes, traumatic events generally involve threats to life or bodily integrity, or a close personal encounter with violence and death. They confront human beings with the extremities of helplessness and terror, and evoke the responses of catastrophe (p. 33).*

That traumatic events are relatively common has been well-documented in a number of recent investigations (e.g., Breslau, Davis, Andreski, & Peterson, 1991; Norris, 1992; Kessler, Sonnega, Bromet, Highes, & Nelson, 1995; Resnick, Kilpatrick, Dansky, Saunders, & Best, 1993; Vrana & Lauterbach, 1994). In the National Comorbidity Study, for example, Kessler et al. (1995) found that 60.7 percent of men and 51.2 percent of women experienced at least one traumatic event in their lives, and that 34.2 percent of men and 24.9 percent of women experienced two or more such events.

In response to these concerns, a two-part definition for Criterion A was developed for the DSM-IV, representing a sharp departure from previous definitions. The first part of the new Criterion A, which requires that "the person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others," addresses the core aspect distinguishing traumatic events from ordinary stressors. The second part, which requires that "the person's response involved intense fear, helplessness, or horror," recognizes that an individual's subjective response to an event is an important dimension to consider in defining a traumatic event.

Given its seemingly greater explicitness, the DSM-IV version of Criterion A can be seen as a more conservative definition of a trauma that should result in fewer experiences being classified as traumatic. However, this definition contains ambiguities that permit considerable interpretive flexibility. For example, while the words "experienced" and "witnessed"

in the first part of the definition seem clear, the phrase "confronted with" could be construed quite broadly. Similarly, the words "death" and "serious injury" seem clear, but the phrase "threat to physical integrity" is quite ambiguous.

The practical assessment problem, then, is how to determine whether or not someone experienced at least one traumatic event that would make them eligible for a PTSD diagnosis. In extreme cases, there is little ambiguity and the decision is straightforward. Prototypical events such as rape, combat, and natural disasters are obviously catastrophic and engender nearly universal fear and horror. But many events are less clear-cut, and for these events the way in which "traumatic" is defined becomes crucial.

For example, a stressor that has received an increasing amount of attention in the last few years is chronic illness, especially cancer (see Green, Epstein, Krupnick, & Rowland, 1997). In the DSM-III, chronic illness was offered as an explicit example of a common stressor that would *not* be considered traumatic. In the DSM-IV, though, at least some chronic illnesses (those considered to be life-threatening) qualify as potentially traumatic stressors. Nonetheless, as Green et al. (1997) have observed, even life-threatening illness is not necessarily a good fit to the DSM-IV definition of a trauma, and this category of stressors remains somewhat of a gray area. Our view is that until more research is available, the most reasonable approach at this time is to attempt to apply Criterion A as it is defined in the DSM-IV, recognizing that some latitude will be needed in order to appropriately classify events that are not prototypical traumatic stressors.

Measuring Traumatic Events. Apart from the difficulties involved in defining traumatic events, there are also problems with assessing them. Until recently, far less attention has been paid to the assessment of trauma than to the assessment of PTSD symptoms (see Krinsley & Weathers, 1995). As a result, there is a paucity of well-validated, standardized measures of trauma, but an abundance of study-specific measures, some of them inadequately conceived and poorly constructed. Very few trauma measures have been used in more than a handful of investigations, and even the best measures have only limited psychometric data supporting their use. Currently, the most well-validated measures are those designed to assess exposure to a single type of stressor, such as combat or rape, but many of these measures are not designed to assess the two-part DSM-IV definition of Criterion A.

There are several practical considerations in assessing traumatic events. First, any PTSD assessment requires some means of establishing that the respondent experienced at least one Criterion A event, whether it be a standardized trauma measure, unstructured interview, or some other approach. If the assessment is focused on a single index event or single type of stressor, it may be sufficient to inquire about the index event to determine whether it meets both parts of Criterion A. In addition, severity of exposure to the index event can be quantified by administering a focused trauma scale. Second, there is a growing recognition that traumatic events are multidimensional and that certain event characteristics (e.g., perceived life threat, degree of actual physical harm, age at onset, chronicity, whether the trauma was interpersonal in nature) may be especially predictive of PTSD symptom severity (see Green, 1993; Resnick et al., 1991). Thus, to merely establish that a traumatic event occurred is to ignore potentially valuable information that could lead to a richer understanding of an individual's unique adaptation to that event.

Third, trauma measures vary widely in scope and format, ranging from self-report checklists assessing the presence or absence of a limited range of potentially traumatic events to comprehensive protocols assessing a wide range of stressors through both self-report and interview. Although the use of checklists and questionnaires is widespread, we are not sanguine that traumatic events can be validly determined solely on the basis of self-report. In our research and clinical assessments we use both checklist and interview formats and often find considerable divergence between them. Discrepancies occur in both directions. Sometimes respondents endorse items on a checklist that an interviewer then judges to be relatively trivial, but other times an interviewer elicits reports of unmistakably traumatic events that were not endorsed on a checklist. Although trauma interviews

are typically more time-consuming and place a greater burden on both respondents and interviewers, they provide the best means for determining that reported events meet the Criterion A definition of a trauma.

In our own work, we use several different trauma measures depending on the nature of the assessment task. For quantifying degree of combat exposure, we use the Combat Exposure Scale (CES; Keane et al., 1989), a 7-item self-report measure. When screening for potential traumatic events to serve as the basis of a PTSD symptom inquiry, we use the Life Events Checklist (LEC), a self-report screening measure that accompanies the DSM-IV version of the CAPS. Positively endorsed items on the LEC, which assesses 17 different stressor categories, are followed by an interview before beginning the PTSD symptom inquiry, which is based on as many as three different traumas. When a history of trauma across the life span is the focus of an assessment, we use the Evaluation of Lifetime Stressors (ELS; Krinsley, 1996), a comprehensive protocol for assessing lifetime trauma that consists of a screening questionnaire and a follow-up interview.

Linking PTSD Symptoms to Specific Traumatic Events. Another issue closely related to the definition and measurement of traumatic events is establishing that PTSD symptoms are attributable to a specific event. Since it was first officially recognized, PTSD has been conceptualized as an essentially unitary syndrome representing an unequivocal change from a previous level of functioning. In the prototypical scenario of a single, relatively circumscribed traumatic event in an adult with no previous history of trauma or psychopathology, it is relatively straightforward to link symptoms to the trauma. But if any aspect of this scenario is changed, the picture becomes muddied. If the clinical presentation involves multiple traumatic events, childhood onset of trauma, prior psychopathology, or an underlying but previously unexpressed vulnerability to psychopathology, the task of linking symptoms to a specific event is greatly complicated.

In the DSM-IV, the first eight PTSD symptoms—including the five reexperiencing symptoms, the two effortful avoidance symptoms, and the amnesia symptom—are inherently linked to the trauma. If a person is having nightmares or intrusive thoughts, what is the content of these intrusions? If they are avoiding internal or external triggers, what event do these triggers remind them of? On the other hand, the remaining symptoms, including the emotional numbing and hyperarousal symptoms, are not explicitly linked to a trauma and therefore require clinical judgment to determine whether they can be attributed to a specific event. The consensus conference recommended that assessors should try to link these remaining symptoms to a specific event if possible, but should count symptoms toward a PTSD diagnosis if they clearly fit phenomenologically, even if they cannot be unequivocally linked to a specific trauma.

To address this issue in our own clinical and research assessments, we have included a *trauma-related* rating on the DSM-IV version of the CAPS for the numbing and hyperarousal symptoms, whereby the link between a symptom and a specific traumatic event is rated as *definite, probable,* or *unlikely.* If a symptom meets the diagnostic criterion phenomenologically, the interviewer asks about the onset of the problem. If there is an unequivocal change from a previous level of functioning following a specific traumatic event, the symptom is coded as definitely trauma-related. If the link is not explicit, but there is compelling evidence that the symptom is functionally related in some way to a traumatic event, it is coded as probably trauma-related.

An example of a *probable* rating is a woman with a history of chronic sexual abuse in childhood who complains of detachment or estrangement from others and reports that she has "always been that way." Another example is a combat veteran who can't state when his restricted range of affect began, but reports that it worsens around the anniversary of a traumatic experience in combat. Here the inference is that the restricted range of affect is functionally related to heightened reexperiencing, and thus to the trauma itself. Symptoms that are coded as either definitely or probably trauma-related are counted toward a PTSD diagnosis. Only when a symptom is clearly attributable to some other cause, with no apparent connection to a specific trauma, would it not be counted toward a diagnosis.

Handling Multiple Traumas. The problem of linking PTSD symptoms to specific events is compounded when individuals report experiencing multiple traumatic events over their lives. A discussion of the cumulative effect of repeated trauma is beyond the scope of this chapter, but its potential complicating effect on the assessment task can be illustrated by considering some of the numerous possible outcomes of experiencing two traumatic events:

1. An initial trauma leads to no PTSD, subthreshold PTSD, or full-blown PTSD.
2. The person remains asymptomatic, recovers fully, recovers partially, remains symptomatic, or experiences a delayed onset of symptoms.
3. Then the person experiences a second trauma which (a) initiates a new PTSD syndrome that would have developed even in the absence of the first event, (b) initiates a new syndrome potentiated by the earlier trauma although no PTSD developed previously, (c) reactivates an old syndrome, (d) exacerbates a current subthreshold syndrome into a full-blown syndrome, or (e) exacerbates a current syndrome into a more severe one.

These possible pathways to a current PTSD symptom picture expand exponentially as the number of traumas increases, making the attribution of symptoms to specific traumas extraordinarily difficult. Given the high prevalence of exposure to multiple traumas that was noted earlier, confronting this issue is more the exception than the rule in the assessment of PTSD. In the hundreds of clinical and research assessments conducted in our lab, the modal presentation is a complicated clinical picture involving multiple traumas across the life span, interacting in complex ways with constitutional factors, personality factors, social learning factors, and environmental factors (see Green, Wilson, & Lindy, 1985). To ask which trauma caused the PTSD symptoms may be reductionistic and inappropriate given this multivariate, nonlinear perspective.

Comorbidity

One of the best-replicated findings regarding the descriptive psychopathology of PTSD is its association with high rates of comorbid disorders, especially depression, substance abuse, and other anxiety disorders (see Keane & Kaloupek, 1997, for a recent review). From a conceptual perspective, this comorbidity is so prevalent that it challenges the notion of PTSD as a distinct diagnostic entity and raises questions about the functional relationships among PTSD, other disorders, and traumatic events. It is unclear if comorbid disorders are best conceptualized as independent psychopathological processes, as integral and normative aspects of the posttraumatic response, or as consequences of chronic PTSD. Comorbidity also may reflect a more general vulnerability to psychopathology that renders some individuals more susceptible to developing a variety of disorders, including PTSD.

From a practical assessment perspective, comorbid disorders complicate differential diagnosis and treatment planning. Diagnostically the task is to elicit information about all of the symptoms with which an individual presents, attribute them as accurately as possible to distinct if overlapping syndromes, and to determine, if possible, which syndrome is primary. The greater the comorbidity, the greater the number of symptoms and the more difficult the task becomes, particularly when there is direct overlap in the criteria for two syndromes. For example, PTSD and depression share anhedonia, sleep disturbance, and impaired concentration as core diagnostic criteria and also arguably overlap in terms of restricted range of affect, impairment in interpersonal functioning, and guilt. Interestingly, although the PTSD section in the DSM-IV discusses differential diagnosis, there are no explicit exclusion criteria for PTSD as there are for other Axis I disorders, including the close counterpart of PTSD, acute stress disorder.

In treatment planning, the task is to identify targets for change in therapy and to develop a strategy for tackling multiple targets hierarchically. The presence of comorbid psychopathology means that a person is presenting with a variety of clinically significant problems, each representing an important target for intervention. Considering the entire clinical picture and weighing all potential benefits and risks, a clinician must decide which problems to address first, and with which interventions. For example, an individual with PTSD who has an extensive history of alcohol and drug abuse, and who is recently sober and struggling against relapse, is a poor candidate for a trauma-focused intervention such as flooding, since

the painful affects that emerge may lead the individual to resort to alcohol or drugs to alleviate their distress.

Closely related to the issue of comorbidity is the growing recognition that certain types of traumatic stress may have a far more pervasive and damaging impact on affected individuals than the current diagnostic criteria for PTSD would suggest. Investigators who have focused on chronic, interpersonal trauma, especially with a childhood onset such as incest, physical abuse, and severe emotional abuse and neglect, argue that the characteristic responses to such stressors involve a much broader range of symptoms, known as complex PTSD (Herman, 1992) or Disorder of Extreme Stress Not Otherwise Specified (DESNOS; see Herman, 1992). According to Herman (1992), the prominent symptoms of this syndrome include affect dysregulation, dissociation, characteristic distortions of personality, markedly impaired interpersonal relationships, and alterations of meaning.

Increasing evidence of variability in posttraumatic syndromes has led some to argue that PTSD is just one of a group of stress-related disorders that includes adjustment disorders, acute stress disorder, PTSD, and complex PTSD (see Brett, 1993; Davidson & Foa, 1991). Such a category exists in the International Classification of Diseases and Related Health Problems (ICD-10; World Health Organization, 1990). In the DSM-IV, however, PTSD and acute stress disorder are classified as anxiety disorders, and the symptoms of complex PTSD are listed as associated features of PTSD.

In sum, traumatized individuals typically present with a variety of clinically significant problems that extend well beyond the diagnostic criteria for PTSD. Therefore, in most assessment contexts a thorough assessment of comorbid psychopathology is essential. Those who assess victims of chronic interpersonal trauma need to be particularly aware that the modal presentation may include many problems other than the core symptoms of PTSD.

Assessing Response Bias

A potentially significant threat to the psychological assessment of PTSD is the problem of response bias, which refers broadly to test-taking behaviors that threaten the validity of inferences made on the basis of test scores. Response bias can take a number of forms, including socially desirable responding or "faking good," malingering or "faking bad," yea-saying, nay-saying, or careless or random responding. In the assessment of PTSD the concern most often centers on malingering or faking bad. Although it is not unique among psychiatric disorders in this regard, PTSD may be particularly susceptible to symptom exaggeration. It is a highly compensable disorder, not only for combat veterans but also in civilian litigation. In addition, it has been used as the basis of an insanity defense and as a mitigating factor in criminal proceedings (see Resnick, 1997).

Unfortunately, most PTSD measures are quite face-valid, meaning that the intent of their items is transparent to anyone, including respondents. Therefore, if someone were motivated to shape their answers to create a desired impression, it would be relatively easy to do so. This is particularly true for self-report measures of PTSD, most of which include no means for detecting response bias. It is less of a problem for structured interviews since clinicians can obtain additional information by making behavioral observations during the interview and requiring respondents to supply compelling descriptions of symptoms, and since final ratings are based on clinical judgment and not just on a respondent's report (see chapter 12 for further discussion of response bias).

Although the issue of response bias is of greatest concern in contexts when there is obvious incentive to exaggerate symptoms, some effort should be made to measure it whenever possible. There are several effective strategies that can be followed to help reduce response bias or to detect it when it occurs.

1. Use multiple sources of information to substantiate the real-world impact of reported symptoms, including job history, legal history, treatment history, and the reports of one or more family members, friends, or co-workers. Whenever possible, corroborate the occurrence of the trauma. For combat veterans this would involve at a minimum an inspection of their discharge papers, and preferably an examination of their full military record if available.

2. As noted earlier, if an interview is used to assess PTSD symptoms, interviewers should make careful behavioral observations to note if any of the reported

problems are manifested during the assessment. Aspects of the syndrome that may reasonably be expected to appear include overt signs of emotional distress, dissociation, avoidance of painful topics, irritability, lapses in concentration, hypervigilance, and exaggerated startle. Also, interviewers should evaluate the extent to which respondents volunteer information or merely endorse symptoms after questions are posed. If they don't do so spontaneously, respondents should be urged to elaborate on their symptoms and to supply concrete behavioral examples. If respondents are unable to elaborate, or if their responses seem unnatural or contrived, the validity of their symptom endorsement is undermined.

3. Administer instruments that contain indicators of response bias. The most widely used measure containing such indicators is the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), which provides detailed information about several different types of response bias, including random responding, symptom exaggeration, and defensive responding. Surprisingly, only one trauma-specific measure, the Trauma Symptom Inventory (TSI; Briere, 1995), contains built-in scales to detect response bias.

4. Administer an instrument specifically designed to detect malingering, such as the Structured Interview of Reported Symptoms (SIRS; Rogers Bagby, Dickens, 1992). The SIRS is fairly simple to administer and score and its use is supported by an extensive body of research. One potential drawback is that the SIRS was developed to detect malingering of psychopathology in general and was not specifically designed to detect feigned PTSD, so little is known about its use for that purpose. We are currently administering the SIRS in several research projects to determine its utility for identifying exaggerated or malingered PTSD symptoms. Even if the SIRS does not work well with PTSD, the strategies it employs are excellent and could easily be generalized to content more appropriate for detecting the malingering of PTSD specifically.

Using Multiple Measures

The use of multiple measures has been strongly advocated in the assessment of PTSD (e.g., Keane et al., 1987; Kulka et al., 1991). The basic argument in favor of this approach is that all measures are fallible to some degree, so that combining information from multiple measures, ideally based on multiple methods, should help reduce diagnostic errors. In principle this argument is compelling, but it raises two key questions: What measures should be included in a PTSD battery, and how should they be combined in order to reach the most valid diagnostic decisions?

In her seminal work on test evaluation, Kraemer (1992) provides an extensive discussion of the complex issues involved in answering these questions and delineates statistical techniques for combining measures and constructing optimal test batteries. According to Kraemer, any legitimate test, defined as a test shown to be significantly associated with a clinical diagnosis based on a widely accepted procedure, is appropriate for inclusion in a test battery. She demonstrates how tests can be combined by using "and/or" rules or by summing weighted scores in a prediction equation.

In general, "and" rules (e.g., respondents receive a positive test if they exceed the cutoff on Test 1 *and* Test 2 *and* Test 3) are more restrictive, resulting in fewer false positives but more false negatives. In contrast, "or" rules (e.g., respondents receive a positive test if they exceed the cutoff on Test 1 *or* Test 2 *or* Test 3) are more inclusive, resulting in fewer false negatives but more false positives. The use of and/or rules permits multiple tests to be administered in sequence, which means that only the minimum number of tests needed to satisfy the rule are administered. The use of weighted scores, however, requires that all tests in a battery be administered, which can be much more time-consuming and costly.

Another approach to combining information from multiple sources is to rely on the consensus of a group of experts who meet to discuss each piece of data and determine PTSD diagnostic status on a case-by-case basis. This is a flexible, idiographic approach that is especially useful for clinical decision-making regarding highly ambiguous or complex cases, and we follow it in our own clinical case conferences. The problem with this approach is that it does not yield an empirically derived, well-defined decision rule that can be applied outside of its unique context. Although

not appropriate for every assessment problem, other solutions for dealing with multiple measures include (a) excluding respondents from a study when there is discordance among measures in a battery, (b) assigning differential weights to the various measures, giving precedence to certain measures such as structured interviews, and (c) designating respondents with discordant measures as subthreshold.

To date, very little research has been conducted on the validity of combining multiple PTSD measures. The NVVRS (Kulka et al., 1990) provides perhaps the best and certainly the most extensive example, combining the use of and/or rules, prediction equations, and clinical decision-making based on expert consensus. Additional studies are sorely needed on this vital issue in order to determine which measures, in which combinations, and for which populations, best predict a PTSD diagnosis.

Clarifying Operational Definitions of PTSD Assessment Procedures

The issues discussed so far reflect various aspects of a broader concern, which is the need for much greater specificity in operational definitions of PTSD assessment procedures. Method sections of journal articles too often contain inadequate information regarding the selection, administration, and scoring of PTSD assessment measures. In general, when PTSD diagnostic status appears as a variable in a study, the method section should explicitly describe: (a) which diagnostic criteria were used (e.g., DSM-IV); (b) which instrument, and which version, was administered; (c) who administered the instrument, including credentials, experience assessing PTSD, training and experience with the specific instrument; (d) conditions under which the instrument was administered, including mode of administration (e.g., phone versus in-person, computer versus paper-and-pencil), setting (e.g., take-home versus lab, individual versus group administration), and when in the overall assessment the instrument was administered; (e) how the instrument was scored, and whether alternative scoring rules were explored; (f) how trauma was assessed, including how an index trauma was identified, how multiple traumatic events were handled, and how symptom-event linking was handled; and (g) the inclusion and exclusion criteria for participation (e.g., male, treatment-seeking Vietnam combat veterans with no lifetime psychotic disorder and no current substance use disorder).

One area in particular that we feel needs much more attention from investigators is the specification and justification of scoring rules for converting dimensional scores to dichotomous diagnoses. Increasingly, the trend in PTSD assessment is toward measures that yield a range of scores for each symptom, indicating gradations of severity of a symptom rather than simply its presence or absence. The dimensional scores that result are ideal for many applications, but when a PTSD diagnosis is needed these dimensional measures must somehow be converted to a dichotomous decision for each individual being assessed.

There are a great many ways in which dimensional PTSD measures could be scored to yield a diagnosis. For most such measures, however, only a single, rationally derived scoring rule has been explored, typically with little empirical evidence supporting its use over alternate rules. This is unfortunate because the choice of a scoring rule can significantly influence the prevalence or base rate of PTSD in a sample. Scoring rules can range from lenient, whereby many individuals receive a diagnosis, to stringent, whereby few individuals receive a diagnosis. In our work with the CAPS, for example, we have developed and evaluated a number of different scoring rules, including rationally and empirically derived rules (Weathers, Meron, & Keane, 1997). We and others (Blanchard et al., 1995) have found that the original CAPS scoring rule, whereby symptoms are coded *present* if CAPS items are rated as *1* or higher for frequency and *2* or higher for intensity (on 4-point scales), is a relatively liberal rule that may "overdiagnose" in some circumstances. For some applications a more conservative rule may be better suited.

Given the lack of empirical evidence to-date on this crucial topic, we feel that it is incumbent upon developers of dimensional PTSD instruments to propose alternative scoring rules and conduct research comparing their diagnostic utility. Those who use these measures should, at a minimum, report which rule they employed in a given context and justify their

choice. Further, when appropriate, they should use several different scoring rules and describe their relative impact on the results of the study.

PSYCHOMETRIC CONSIDERATIONS

In this section we provide a brief overview of the psychometric concepts, procedures, and issues involved in developing and evaluating new questionnaires or interviews (for a full treatment of this topic, see Weathers, Keane, King, & King, 1997). Psychological assessment instruments are evaluated in terms of their reliability and validity. Reliability is the degree to which test scores are free of measurement error, whereas validity is the degree to which test scores actually reflect the characteristic the test is thought to assess. In general, a reliable test yields consistent scores over repeated administrations and is relatively unaffected by potential sources of error, such as items and testing occasions. For PTSD instruments, as with other measures of psychopathology, the types of reliability that are of greatest concern are *internal consistency,* which refers to consistency over items on a test; *test-retest reliability,* which refers to consistency over repeated administrations of a test; and *interrater reliability,* which refers to consistency in test scores over different raters.

Internal consistency can be evaluated following a single administration of a test. It is typically quantified and reported as an alpha coefficient, an index that ranges from 0.00 to 1.00, with scores closer to 1.00 indicating greater intercorrelation of items. It is often accompanied by a report of the correlations between individual items and the total score for the remaining items (*item-scale total correlations*), which indicate how well individual items fit the other items on the scale. Test-retest reliability is determined by administering an instrument on two occasions and calculating the correlation between the two scores. One complication in a test-retest design is the length of the interval between administrations. If too brief a period is used (say, an hour), respondents' memory for how they answered previously may influence the second administration, so memory may be confounded with consistency of responding. On the other hand, if too long an interval is used (say, a month), genuine change in clinical status may occur, confounding cli ical change with consistency of responding.

For interviews, another pertinent type of reliabi ity is interrater reliability, which involves two adc tional sources of error: differences in ho interviewers elicit information and differences in ho they score responses. In a simple interrater study, tw or more raters observe the same interview and mal simultaneous ratings. Alternatively, the interview c be audiotaped or videotaped and rated at a later tim This answers the question of whether different rate can make comparable ratings when they are expose to the same information. A more stringent test of rel ability is to have two raters conduct separate inte views on different occasions and make independe ratings. Like the simple interrater design, this "secor opinion" design considers errors due to the way di ferent raters use the scale, but also considers erro due to the way that different raters elicit informatio For instruments yielding dimensional scores, inte rater reliability is reported as a correlation coefficie or an intraclass correlation coefficient. For instr ments yielding dichotomous (present/absent) dec sions, reliability is reported as a kappa coefficient, measure of agreement that is corrected for chanc agreement between raters. A kappa of 0.00 indicat only chance agreement between two raters, whereas kappa of 1.00 indicates perfect agreement.

Reliable instruments, though, are not necessaril valid. Validity refers to the accumulation of evidenc supporting the inferences, interpretations, or concl sions that are made on the basis of test results. Thre types of validity are considered. *Content validity* r fers to evidence that items on an instrument ad quately reflect the domain of interest. The evaluatio of content validity is a complex, multistage proces that typically includes such steps as specifying th main purposes of the instrument, carefully definin the construct, determining item format, generating pool of items, and reviewing and revising items (se Haynes, Richard, & Kubany, 1995). Expert judgme plays a crucial role throughout the content validatio process, particularly in the initial stages in which th instrument is conceived and items are created and r vised. Content validation is essential for rationall developed instruments, such as the Mississippi Scal

which are developed in accordance with *a priori* theoretical conceptualizations. It is much less of a consideration for empirically developed measures, such as the PK scale of the MMPI and MMPI-2 (Keane, Malloy, & Fairbank, 1984; Lyons & Keane, 1992), which consist of items selected solely on the basis of their ability to discriminate individuals with and without a disorder.

Criterion-related validity refers to evidence that the test predicts some variable of interest, such as an outcome, behavior, or diagnosis. When the test and the criterion are measured at approximately the same time, this is known as *concurrent validity.* If the criterion is measured at some point after the test, this is known as *predictive validity.* A frequently evaluated form of concurrent validity for PTSD instruments is *diagnostic utility,* which is the extent to which an instrument predicts a PTSD diagnosis. Indices of diagnostic utility include *sensitivity* (the proportion of those with a positive diagnosis who have a positive test), *specificity* (the proportion of those with a negative diagnosis who have a negative test), and *efficiency* (the proportion of overall agreement between test and diagnosis). For PTSD measures that yield continuous scores, the diagnostic utility of a range of cutoff scores can be determined, and different cutoffs can be employed for different assessment purposes. In general, there is a tradeoff between sensitivity and specificity. Lenient cutoffs have greater sensitivity (but lower specificity) and are appropriate for screening, stringent cutoffs have greater specificity (but lower sensitivity) and are appropriate for confirming a diagnosis, and moderate cutoffs balance sensitivity and specificity and are appropriate for differential diagnosis.

Construct validity is the broadest form of validity and can be seen as subsuming the other types of validity. It refers to evidence that a test is primarily a measure of the construct of interest and not of other constructs. One of the most common approaches to establishing construct validity is to examine the pattern of correlations in a multitrait-multimethod correlation matrix (Campbell & Fiske, 1959), in which multiple measures of the construct of interest are compared to multiple measures of other conceptually distinct constructs. The construct validity of a measure is demonstrated by showing that it correlates strongly with other measures of the same construct (*convergent validity*) and correlates weakly with measures of other constructs (*discriminant validity*).

A final consideration regarding validity is what is known as *face validity.* Despite its name, this is not a true form of validity in the same sense as the other three types previously described. Rather, it refers to the degree to which the intent of a test and the content of its items are obvious to anyone who reads it, including test takers. As noted earlier, face validity renders many PTSD assessment instruments susceptible to symptom exaggeration or other types of response bias. Nonetheless, face validity may be unavoidable for rationally derived instruments and may even be desirable. When content validation procedures are carefully implemented, test items *should* reflect the domain of interest, and thus their content and intent will be obvious even to untrained observers. Another factor to consider in this regard is consumer satisfaction: Individuals undergoing an assessment for PTSD may feel puzzled or annoyed if test items appear to be irrelevant, even if the test is valid for assessing PTSD.

PTSD ASSESSMENT INSTRUMENTS

In this section we briefly review a number of structured interviews and self-report instruments for assessing PTSD, including well-established and widely used measures as well as several promising, newly published measures (for further description and contact information regarding these and other measures, see Stamm, 1996). When selecting a PTSD instrument clinicians or investigators must first address the fundamental validity question: What task is the instrument intended to accomplish and what evidence supports its use for that purpose? All of the conceptual, methodological, and psychometric issues discussed earlier are relevant to this question and will shape an appropriate answer. The next step is to address a number of practical considerations, including the amount of time available for the assessment, whether interviewers are available, and whether the format and content of the instrument are appropriate for the assessment task (see Weathers, Keane et al., 1997).

PTSD measures vary widely in format, most importantly in terms of (a) the method of administration (i.e., interview, paper and pencil, computer); (b) the wording of items or prompt questions; (c) the nature of the rating scale, including both the number of points on the scale and the way in which symptom severity is defined in the rating scale anchors (e.g, in terms of frequency of symptoms or subjective distress); and (d) the time frame assessed (e.g., past week, past month). In terms of content, all structured interviews are based directly on the DSM diagnostic criteria. However, the PTSD consensus conference divided the self-report measures of PTSD into three categories: those corresponding exactly to the DSM criteria, those tapping the core and associated features of PTSD but not corresponding exactly to the DSM criteria, and those derived empirically from existing instruments.

The consensus conference recommended that a structured interview be used to assess PTSD whenever feasible. It further recommended that a structured interview, administered by an experienced clinician and yielding both continuous and dichotomous scores, be required when an in-depth assessment of PTSD is the focus of an investigation. According to the consensus conference, DSM-correspondent self-report measures are appropriate for screening, for survey research where interviewing is not possible, or as adjuncts to a structured interview. In contrast, measures that are PTSD-focused but not DSM-correspondent should be administered only in conjunction with a structured interview or a DSM-correspondent self-report measure whenever possible. Empirically derived instruments should also be administered in conjunction with other measures, and should only be used as the sole measure of PTSD when no other measures are available, as in the case of archival data sets.

Interviews

Structured Clinical Interview for DSM-III-R (SCID). The SCID (Spitzer et al., 1990) is a comprehensive structured interview that assesses all of the major Axis I disorders. Although the published version of the SCID for the DSM-III-R did not contain a PTSD module, one was created for use in the NVVRS (Kulka et al., 1990) and has since become one of the most widely used PTSD interviews. A revised PTSD module will be included in the DSM-IV version of the SCID. As with other modules of the SCID, the PTSD module follows the DSM diagnostic criteria exactly. A standard prompt question is provided for each of the 17 PTSD symptoms, and interviewers rate each symptom as *? = inadequate information, 1 = absent, 2 = subthreshold,* or *3 = threshold.* Interviewers are encouraged to ask additional questions or to clarify as needed in order to obtain accurate information.

The SCID PTSD module appears to have adequate reliability and validity. In the NVVRS, Kulka et al. (1991) found a kappa of .93 when audiotaped SCID interviews were independently scored by a second clinician. In a more stringent test of reliability, Keane et al. (1997) found a kappa of .68 when the SCID PTSD module was administered twice by independent clinicians. Also, in the NVVRS the SCID was positively associated with other measures of PTSD, including the Mississippi Scale (kappa = .53), and the PK scale of the MMPI (kappa = .48), and had excellent sensitivity (.81) and specificity (.98) against a composite PTSD diagnosis (Kulka et al., 1991; Schlenger et al., 1992). The primary limitation of the SCID is that it yields essentially dichotomous data at the item and syndrome levels and thus is not well-suited for quantifying or detecting changes in symptom severity.

Diagnostic Interview Schedule (DIS). The DIS is a highly structured, comprehensive interview designed for use by lay interviewers in the context of epidemiological research. Variants of the DIS PTSD module have been used in all of the recent major epidemiological investigations of PTSD (Breslau, Davis, Andreski, & Peterson, 1991; Centers for Disease Control, 1988; Helzer, Robins, & McEvoy, 1987; Kessler et al., 1995; Kulka et al., 1990; Resnick et al., 1993). Like the SCID, the DIS provides a standard prompt question for each of the 17 PTSD symptoms, but interviewers are discouraged from clarifying any of the standard questions or making additional inquiries. Each symptom is scored dichotomously to indicate its presence or absence.

Although the DIS has been widely used, serious concerns have been raised about its diagnostic utility. In the preliminary validation study of the NVVRS, the DIS was found to be one of the best predictors of a clinical diagnosis of PTSD, with a sensitivity of .87, a specificity of .73, an efficiency of .84, and a kappa of .64. However, in the clinical examination component, when tested against a SCID-based PTSD diagnosis, specificity was quite high (.98), but sensitivity (.22) and kappa (.26) were poor (Kulka et al., 1991). Other investigators have subsequently made a number of modifications to the DIS module to improve its diagnostic utility, and these efforts appear to have been reasonably successful (e.g., Kessler et al., 1995; Resnick et al., 1993).

PTSD Symptom Scale—Interview (PSS-I). The PSS-I (Foa, Riggs, Dancu, & Rothbaum, 1993) is a structured interview specifically designed to assess DSM-III-R PTSD symptoms. It contains 17 items, each consisting of a single prompt question corresponding to one of the 17 diagnostic criteria for PTSD. Interviewers rate the severity of each symptom over the past two weeks as *0 = not at all, 1 = a little bit, 2 = somewhat,* or *3 = very much.* A total severity score is obtained by summing ratings over all 17 items. A PTSD diagnosis is obtained by considering symptom ratings of *1* or higher as *present,* then following the DSM-III-R diagnostic algorithm.

The PSS-I appears to have excellent psychometric properties. Foa et al. (1993) reported an alpha coefficient of .85 for all 17 items and an average item-scale total correlation of .45. Test-retest reliability for the total severity score was .80, and the kappa coefficient for a diagnosis of PTSD was .91. Against a SCID-based PTSD diagnosis, the PSS-I had a sensitivity of .88, a specificity of .96, and an efficiency of .94. The PSS-I also correlated strongly with the IES and the Rape Aftermath Symptom Test (RAST; Kilpatrick, 1988), as well as the BDI and STAI.

The advantages of the PSS-I are that it yields continuous and dichotomous scores, is easy to administer, and has good reliability and validity for assessing PTSD. The disadvantages are that it includes only a single prompt question for each item, its rating anchors are not explicitly defined, it assesses symptoms over a two-week rather than the one-month period required in the DSM-IV PTSD criteria, and it does not address lifetime diagnostic status. Also, the proposed scoring rule was rationally rather than empirically derived, and alternative rules have not been explored.

Structured Interview for PTSD (SI-PTSD). The SI-PTSD (Davidson, Smith, & Kudler, 1989) is a structured interview originally designed to assess both DSM-III and DSM-III-R criteria for PTSD. Items consist of initial prompt questions and follow-up questions that clarify the initial question with concrete behavioral examples. The severity of each symptom is rated on a 5-point scale, both for the past month and for the worst period since the trauma. Rating scale anchors vary across items, but generally follow the pattern of *0 = none, 1 = mild, 2 = moderate, 3= severe, 4 = extremely severe.* Descriptors are provided for rating scale anchors to clarify what is meant by a given rating. A total severity score is obtained by summing ratings over all 17 symptoms, and symptoms are counted toward a PTSD diagnosis when they are rated as *2* or higher.

The SI-PTSD appears to be psychometrically sound. Davidson et al. (1989) reported an overall alpha of .94, test-retest reliability of .71, and remarkably strong interrater reliability, with intraclass correlations ranging from .97 to .99 and 100 percent diagnostic agreement. Against a SCID diagnosis of PTSD, the SI-PTSD had a sensitivity of .96, a specificity of .80, and a kappa of .79.

The advantages of the SI-PTSD are that it yields both continuous and dichotomous scores, it provides follow-up prompts and descriptors for rating scale anchors, and it has documented reliability and validity. The disadvantages are that it relies on a single, rationally derived scoring rule, and that it uses a "worst ever" convention for assessing lifetime ratings for individual symptoms, making it difficult to establish that lifetime symptoms occurred as part of a syndrome.

PTSD Interview (PTSD-I). The PTSD-I (Watson, Juba, Manifold, Kucala, & Anderson, 1991) is a structured interview for assessing the DSM-III-R criteria for PTSD. Individual PTSD symptoms are inquired with a single prompt question, and severity

ratings are made on a 7-point scale ranging from *1 = No/Never* to *7 = Extremely/Always*. A total severity score is computed by summing ratings over the 17 symptoms, and symptoms are considered present if they are rated as *4 = Somewhat/Commonly* or higher. Lifetime diagnostic status is established through several questions about chronology that follow symptom inquiry.

Watson et al. (1991) reported strong reliability and validity for the PTSD-I. The alpha coefficient over all 17 items was .92. Test-retest reliability over a 1-week interval was .95, with 87 percent diagnostic agreement between the two administrations. Against a PTSD diagnosis based on the DIS, the PTSD-I had a sensitivity of .89, a specificity of .94, an efficiency of .92, and a kappa of .84.

The PTSD-I yields both continuous and dichotomous scores and appears to have desirable psychometric properties. A major limitation, however, is the recommended format for administration. Interviewers are instructed to give respondents a copy of the rating scale, read the questions aloud, and ask respondents to rate themselves. Like the DIS, little or no clinical judgment is involved in the assessment of PTSD symptoms. Although the PTSD-I could be implemented in other ways, the current format is more like a questionnaire than a structured interview.

Clinician-Administered PTSD Scale (CAPS). The CAPS (Blake et al., 1990; Blake et al., 1995) is a comprehensive structured interview for PTSD developed at the National Center for PTSD. Intended for use by clinicians experienced with PTSD and structured interviewing, the CAPS has several features that were designed to address some of the limitations of other PTSD interviews. First, the CAPS assesses the 17 core symptoms of PTSD, as well as associated symptoms, response validity, overall symptom severity, and the impact of symptoms on social and occupational functioning. Second, the CAPS assesses the frequency and intensity of each symptom on separate 5-point rating scales, yielding continuous and dichotomous scores for each symptom and across all 17 symptoms. Third, the CAPS contains behaviorally anchored prompt questions and rating scales to help increase the reliability of symptom inquiry and severity ratings. Fourth, the CAPS provides specific guidelines for assessing lifetime PTSD diagnostic status, ensuring that a respondent experienced symptoms as part of a syndrome lasting at least one month. Finally, several rationally and empirically derived scoring rules have been developed for converting continuous frequency and intensity scores into dichotomous scores (Weathers, Meron, & Keane, 1997).

As with most of the other interviews described in this section, the CAPS has very strong reliability and validity. Weathers, Blake et al. (1997) found an alpha coefficient of .94 for the 17 core PTSD symptoms, and test-retest reliability coefficients ranging from .90 to .98 for total severity. Using the optimal empirically derived scoring rule, the test-retest reliability for a PTSD diagnosis was .89. Against a PTSD diagnosis based on the SCID, this same scoring rule had a sensitivity of .91, a specificity of .86, an efficiency of .89, and a kappa of .77. With respect to convergent and discriminant validity, the CAPS correlated strongly with other measures of PTSD, including the Mississippi Scale and the PK scale, and correlated moderately with measures of anxiety and depression, but correlated only weakly with measures of antisocial personality disorder.

The CAPS was recently revised for the DSM-IV. Notable changes include (a) adding a protocol to assess Criterion A, including a screening questionnaire and follow-up prompts administered by the interviewer; (b) adding a *trauma-related* convention that requires interviewers to rate the link between a symptom and a specific trauma as *definite, probable,* or *unlikely;* (c) replacing several associated features with items assessing the dissociative symptoms that are included in the criteria for acute stress disorder; (d) clarifying the rating scale anchors to enhance consistency of ratings across items. Also, to avoid confusion about the two forms of the CAPS, the current and lifetime diagnostic version, previously known as the CAPS-1, is now called the CAPS-DX, and the one-week symptom status version, previously known as the CAPS-2, is now called the CAPS-SX. Although the CAPS has many strengths, its major limitation is that it typically takes longer than other PTSD interviews to administer.

Self-Report Measures

DSM-Correspondent Measures.

PTSD Checklist (PCL). The PCL (Weathers, Litz, Herman, Huska, & Keane, 1993; Weathers, Litz et al., 1997) is a 17-item PTSD scale developed in 1990 at the National Center for PTSD for use in establishing the construct validity of the CAPS. Originally based on DSM-III-R PTSD criteria, it was revised in 1994 to correspond to the DSM-IV PTSD criteria. Using a 5-point scale ranging from *1 = not at all* to *5 = extremely,* respondents rate how much they were bothered by each PTSD symptom over the past month. There are three versions of the PCL, which differ only in the target event specified in the reexperiencing and effortful avoidance items. The civilian version (PCL-C) refers broadly to *a stressful experience from the past,* the military version (PCL-M) refers to *a stressful military experience,* and the specific version (PCL-S) refers to a specific stressor identified by the respondent. The PCL-C and the PCL-M are appropriate when a specific trauma has not been identified, and the PCL-S is appropriate when a trauma history has been taken and a stressor has been targeted for symptom inquiry.

Weathers, Litz et al. (1997) reported excellent psychometric properties for the PCL across four different samples of male and female veterans. In the original validation sample of 123 male Vietnam veterans, test-retest reliability was .96 and alpha over all 17 items was .97. The PCL correlated strongly with other measures of PTSD, including the Mississippi Scale, the PK scale, and the IES, and correlated moderately with level of combat exposure. Against a PTSD diagnosis based on the SCID, a cutoff score of 50 on the PCL yielded a sensitivity of .82, a specificity of .83, and a kappa of .64. The PCL has been used successfully in other traumatized populations as well, including motor vehicle accident victims (Blanchard, Jones-Alexander, Buckley, & Forneris, 1996) and breast cancer survivors (Cordova et al., 1995).

PTSD Symptom Scale-Self-Report (PSS-SR). The PSS-SR (Foa et al., 1993) is the self-report counterpart to the PSS-I. With the exception of six reworded items, the PSS-SR is identical to the PSS-I in item content, rating scale, and scoring. Like the PSS-I, the PSS-SR has solid reliability and validity. According to Foa et al. (1993), test-retest reliability for total severity over a 1-month interval was .74, and alpha for the 17 items was .91. with respect to validity, the PSS-SR correlated strongly with the RAST, the IES, the BDI, and the STAI. It also had a sensitivity of .62, a specificity of 1.00, and an efficiency of .86 against a SCID-based PTSD diagnosis. Regarding diagnostic agreement between the PSS-SR and the PSS-I, Foa et al. found a kappa of .73, indicating good correspondence between the two formats.

Falsetti, Resnick, Resick, and Kilpatrick (1993) revised the PSS-SR, slightly rewording six items and adding a 4-point rating scale to assess the frequency of each PTSD symptom. The resulting measure, known as the Modified PTSD Symptom Scale-Self-Report (MPSS-SR), yields both continuous and dichotomous scores and appears to have good psychometric properties. Like the other versions of the PSS, however, the MPSS-SR assesses symptoms over the past two weeks and does not assess lifetime symptom status.

Purdue PTSD Scale-Revised (PPTSD-R). The PPTSD-R (Lauterbach & Vrana, 1996) is a 17-item measure corresponding to the DSM-III-R PTSD criteria. Items are rated on a 5-point scale, from *1 = not at all* to *5 = often,* indicating the frequency of symptoms over the past month. Lauterbach and Vrana (1996) reported excellent psychometric properties for the PPTSD-R in three samples of college students. Test-retest reliability for total severity over a 2-week interval was .72, and alpha over all 17 items was .91. Further, the PPTSD-R correlated strongly with the IES and the civilian version of the Mississippi Scale, and correlated moderately with the STAI and the BDI. Finally, PPTSD-R scores were significantly higher in participants reporting more severe traumatic events, and also were higher in those seeking treatment for trauma-related difficulties relative to those seeking treatment for other reasons as well as non-treatment-seeking participants. The PPTSD-R has promise as a self-report measure of PTSD, although more research is needed with other populations. Research is also needed to determine its diagnostic utility against a PTSD diagnosis based on a structured interview.

Posttraumatic Stress Diagnostic Scale (PDS). The PDS (Foa, 1995) is a newly published self-report measure of PTSD that has the distinction of being the only self-report measure to assess all six (A-F) criteria for PTSD in the DSM-IV. Part 1 of the PDS is a 13-item checklist of potential traumatic events. Part 2 consists of eight items that help determine if an event meets the DSM-IV definition of Criterion A. Part 3 assesses the frequency over the past month of the 17 PTSD symptoms, using a 4-point scale ranging from *0 = Not at all or only one time* to *3 = 5 or more times a week/almost always.* Part 4 assesses the impact of symptoms on various aspects of social and occupational functioning. The PDS yields both a dichotomous diagnostic score and a continuous symptom severity score. An individual PTSD symptom is counted as *present* if the corresponding PDS item is endorsed as a *1* or higher.

The PDS manual describes desirable psychometric properties for the scale in a normative sample of 248 adults assessed in a variety of clinical and research settings. Over an interval of approximately two weeks, test-retest reliability for symptom severity was .83, with a kappa of .74 for diagnostic agreement between the two administrations. The PDS had reasonable diagnostic utility against a PTSD diagnosis based on the SCID, with a sensitivity of .82, a specificity of .77, an efficiency of .79, and a kappa of .59. Convergent validity was demonstrated through strong correlations with the IES, the STAI, and the BDI. Although these results require confirmation in other studies, the PDS appears to have considerable promise as a measure of the DSM-IV PTSD criteria.

Davidson Trauma Scale (DTS). The DTS (Davidson, 1996) is another recently published self-report measure of PTSD, consisting of 17 items corresponding to the DSM-IV PTSD symptoms. Respondents first identify a traumatic event that is most disturbing to them, then rate the frequency and severity of each symptom over the past week on separate 5-point rating scales. Frequency ratings range from *0 = Not at all* to *4 = Every day* and severity ratings range from *0 = Not at all distressing* to *4 = Extremely distressing.* The DTS appears to be designed primarily to yield continuous scores of PTSD symptom severity. The manual provides a novel table for converting total DTS scores into a PTSD diagnosis, based on the ratio of PTSD cases to non-cases at each cutoff score as well as on the expected base rate of PTSD in the target population. However, no scoring rules are provided for converting frequency and severity ratings for individual symptoms into dichotomous scores.

Based on data from four different samples described in the manual, the DTS appears to be psychometrically sound. Test-retest reliability for total DTS scores over a 1-week interval was .86, and alpha coefficients for frequency, severity, and total scores were all greater than .90. The DTS also showed good convergent validity, correlating strongly with the CAPS, the IES, and several other measures of trauma-related psychopathology. Further, the DTS was found to be sensitive both to differences in clinical severity of PTSD and to improvement in symptoms as a result of treatment. Finally, the manual provides diagnostic utility data for several cutoff scores on the DTS. A cutoff of 40, for example, had a sensitivity of .69, a specificity of .95, and an efficiency of 83.

Like the PDS, the DTS shows promise as a self-report measure of PTSD. The one-week time frame is at odds with the DSM-IV requirement of a one-month duration for PTSD symptoms, but it might be effective for tracking treatment progress when multiple assessments over brief intervals are desired. Also, it may be useful to devise and evaluate different scoring rules for the DTS, especially for converting frequency and severity scores to dichotomous scores.

PTSD-Focused, Non-DSM-Correspondent Measures.

Impact of Event Scale (IES). The IES (Horowitz et al., 1979) was the first standardized measure of posttraumatic symptomatology to appear and is also one of the most widely used. Based on Horowitz's (1976) conceptual model of responses to stressful life events, the IES consists of 15 items, including 7 items assessing intrusive symptoms and 8 items assessing avoidance. Respondents first specify a stressful event, then rate the frequency of each symptom over the past week on a 4-point scale, with anchors of *Not at all, Rarely, Sometimes,* and *Often.* Horowitz et

al. (1979) reported strong test-retest reliability (.87 for intrusion and .79 for avoidance) and internal consistency (alphas of .79 for intrusion and .82 for avoidance). Horowitz et al. also found a correlation of .42 between the two subscales, suggesting that intrusion and avoidance are related but distinct responses to significant stressors. These psychometric findings were confirmed by Zilberg, Weiss, and Horowitz (1982), and the distinction between intrusion and avoidance has been supported through several factor analyses (e.g., Schwarzwald, Solomon, Weisenberg, & Mikulincer, 1987; Zilberg et al, 1982).

Although the IES taps two key dimensions of PTSD, it does not correspond completely to the PTSD diagnostic criteria. The IES was recently revised to more directly assess the PTSD symptoms in the DSM-III-R and DSM-IV (Weiss & Marmar, 1997). The most significant change was the addition of six items assessing hyperarousal and one item assessing dissociative reexperiencing. These new items bring the IES closer to the current diagnostic criteria, although the fit is still not as close as it is for the DSM-correspondent measures described in the previous section. The IES does not assess some PTSD diagnostic criteria at all (e.g., diminished interest, foreshortened future), and assesses others only indirectly or ambiguously (e.g., restricted range of affect, inability to recall the trauma).

Other important revisions to the IES include two changes to the rating scale. First, rather than rating symptom frequency, respondents are now instructed to rate the degree of distress associated with each symptom. Second, symptoms are now rated on a 5-point scale, ranging from *0 = Not at all* to *4 = Extremely,* instead of the original 4-point scale. These changes should not preclude reasonable comparability with previous research employing the original IES, but given the confusion over scoring of the original IES (see Green, 1991), test users will need to carefully specify which version of the IES they used and how it was scored.

Mississippi Scale for Combat-Related PTSD (Mississippi Scale). The Mississippi Scale (Keane et al., 1988) is the most widely used measure of combat-related PTSD and has been the subject of extensive, sophisticated psychometric analysis. It consists of 35 items selected from an initial item pool of 200 items based on the DSM-III PTSD criteria and associated features. Items are rated on a 5-point scale, with anchors that vary according to item content (e.g., *1 = Never* to *5 = Very Frequently, 1 = Never True* to *5 = Always True*).

The Mississippi Scale has demonstrated consistently strong psychometric properties across a growing number of investigations. In the NVVRS, it was chosen as one of the primary PTSD indicators based on its excellent performance in the preliminary validation study and the clinical examination component (Kulka et al., 1991). Keane et al. (1988) found an alpha of .94 over all 35 items, and a one-week test-retest reliability of .97. They also found that a cutoff of 107 had a sensitivity of .93, a specificity of .89, and an efficiency of .90 against a consensus clinical diagnosis of PTSD. These results were replicated and extended by McFall et al., (1990), who found moderate to strong correlations between the Mississippi Scale and measures of combat exposure, PTSD, and anger. Recent research has focused on intensive examination of the psychometric properties of individual items (King, King, Fairbank, Schlenger, & Surface, 1993) and on the scale's underlying factor structure (King & King, 1994).

A civilian version of the Mississippi Scale, developed for assessing civilian PTSD in the NVVRS, is also available. The main modification involved rephrasing items containing references to the military. The latest version includes four items added to provide better coverage of the DSM-III-R criteria for PTSD. An initial investigation of its psychometric properties indicated that it performs well but may benefit from some revision (Vreven, Gudanowski, King, & King, 1995).

Penn Inventory for Posttraumatic Stress Disorder (Penn Inventory). The Penn Inventory (Hammarberg, 1992) is a 26-item scale based on the DSM-III and DSM-III-R PTSD criteria. Its item format differs from the other self-report measures described in this section. Modeled after the BDI, items on the Penn consist of four statements, scored 0 to 3, reflecting

increasing symptom severity. In terms of item content, coverage of the DSM-III-R criteria is not complete (e.g., no items explicitly for hypervigilance, physiological reactivity, effortful avoidance), and many items tap symptoms that are not part of the core criteria for PTSD (e.g., alienation, alterations in self-perception, disruptions in goal-directed behavior, grief), although a case could be made that they are important associated features.

In three samples including combat veterans and civilian trauma survivors, Hammarberg reported excellent psychometric properties for the Penn Inventory. In the first sample, across all participants and items alpha was .94 and test-retest reliability was .96. An overall alpha of .94 was replicated in the second sample. A cutoff of 35 yielded a sensitivity of .90, a specificity of 1.00, and an efficiency of .94, although these figures are likely partially attributable to the sampling scheme used to create the study groups. This high level of diagnostic accuracy was also replicated in the second sample. The Penn Inventory correlated strongly with the IES and the Mississippi Scale, as well with the BDI, BAI, and STAI, and correlated moderately with level of combat exposure.

Empirically Derived Measures.

PK Scale of the MMPI and MMPI-2. The original PK scale (Keane et al., 1984) consists of 49 items on the MMPI that were found to statistically discriminate Vietnam combat veterans with and without PTSD. Keane et al. (1984) found that a cutoff score of 30 correctly classified 82 percent of veterans in both a validation and a cross-validation sample. Subsequent investigations of the PK scale have generally confirmed its diagnostic utility, although results have varied considerably across studies, probably largely due to differences in sample characteristics and diagnostic procedures. Also, mean scores and optimal cutoff scores have typically been lower than those reported by Keane et al.

Watson, Kucala, and Manifold (1986), for example, found a sensitivity of .87 and a specificity of .74 when distinguishing between combat veterans with PTSD and normal controls, but a sensitivity of .73 and a specificity of .53 for distinguishing between PTSD patients and patients with other psychiatric disorders. Cannon, Bell, Andrews, and Finkelstein (1987) found a sensitivity of .76 and a specificity of .64 for distinguishing between veteran psychiatric inpatients with and without PTSD. The PK scale has also been used successfully with civilian trauma victims. Koretzky and Peck (1990) found that a cutoff score of 19 correctly classified 87 percent and 88 percent of participants in two different samples.

When the revised version of the MMPI, the MMPI-2, was published, two minor modifications were made to the PK scale, including dropping three repeated items and rewording one item (see Lyons & Keane, 1992). In the normative samples for the MMPI-2, the 46-item PK scale had good internal consistency, with alphas of .85 for males and .87 for females, and good test-retest reliability, with correlations of .86 for males and .89 for females.

The utility of the PK scale in a stand-alone format has also been explored. Lyons and Scotti (1994) found that 94 percent of the veterans in their sample were classified similarly on the PK scale of the MMPI when it was administered both in the context of the full MMPI and as a separate instrument. Herman, Weathers, Litz, & Keane (1996) conducted an comprehensive evaluation of the PK scale of the MMPI-2 as a stand-alone measure. They found a correlation of .90 between the stand-alone and embedded versions of the PK scale. For the stand-alone version, alpha was .96 and test-retest reliability over 2–3 days was .95. A cutoff of 24 on the stand-alone version yielded a sensitivity of .82, a specificity of .76, and a kappa of .59 against a SCID-based PTSD diagnosis. The stand-alone version also correlated strongly with the Mississippi Scale, the IES, the PCL, and the CAPS, and correlated moderately with combat exposure.

Symptom Checklist-90-R (SCL-90-R) PTSD Scales. The SCL-90-R (Derogatis, 1983) has been widely used in PTSD assessment batteries and several investigators have attempted to derive PTSD subscales from its items. For example, Saunders, Arata, and Kilpatrick (1990) identified 28 SCL-90-R items that discriminated female crime victims with and without PTSD. This subscale, the Crime-Related PTSD scale

(CR-PTSD), had an alpha of .93 and correctly classified 89 percent of the sample. Using a similar approach, Weathers et al. (1996) identified 25 items that discriminated combat veterans with and without PTSD. This subscale, the War-Zone-Related PTSD scale (WZ-PTSD), which shares fewer than half its items with the CR-PTSD scale, had an alpha of .97. A cutoff score of 1.3 had a sensitivity of .90, a specificity of .65, and a kappa of .58. This put it second only behind the Mississippi Scale in terms of diagnostic utility, and higher than the Global Severity Index (GSI), although not to a statistically significant degree. These results were replicated in a cross-validation sample. Green (1991) has argued that while the SCL-90-R may be useful in the assessment of PTSD, PTSD subscales are not likely to improve significantly on what can be achieved with the GSI.

ASSESSMENT SCENARIOS

In this section we present a number of assessment scenarios that typify many of our clinical and research endeavors. Reflecting a variety of assessment questions, these scenarios illustrate our solutions to many of the issues raised throughout this chapter. These solutions are intended to be descriptive rather than prescriptive. With respect to the selection of specific measures, we use the instruments with which we are most familiar, either those we have developed in our lab or those we feel are best suited to our needs. As can be seen from the previous section, there are multiple measures in each category that would perform adequately, so our main goal is to encourage others to adopt standardized measures for each of the essential tasks, not to insist that certain measures be used over others. Each assessment situation poses unique challenges requiring tailor-made solutions. However, to the extent that these scenarios generalize to those faced by other clinicians and investigators, they may provide a helpful starting point.

Scenario 1: Brief Clinical Assessment

Our clinical assessments vary according to the referral question, ranging from a relatively brief (1–2 hour) inpatient consultation to a comprehensive (8–10 hour) outpatient evaluation. A typical brief consultation, for which the goal is usually to simply determine if the respondent has a diagnosis of PTSD, consists of the following components:

1. A history of the present illness, including a review of previous episodes, previous treatment, and precipitants of current episode; screening for current and past psychopathology; and a brief review of social history. The overview section of the SCID is a very useful guide for this portion of the evaluation.
2. A CAPS to assess current and lifetime PTSD. This includes the Life Events Checklist (LEC) to screen for lifetime trauma, which is completed by the respondent and followed up by the clinician. Symptom inquiry focuses primarily on the identified trauma, which in our clinic is usually war-zone trauma, but also may focus on other traumas identified by the LEC. To reach a categorical diagnostic decision we use the original, rationally derived scoring rule whereby an item rated with a frequency of "1" or higher and an intensity of "2" or higher counts as a symptom toward a PTSD diagnosis. We then follow the DSM-IV diagnostic algorithm, which calls for at least one reexperiencing symptom, three avoidance and numbing symptoms, and two hyperarousal symptoms, plus the requirements regarding duration and subjective distress or functional impairment. We view this as a minimum threshold, since the "1–2" rule, with the minimum number of symptoms in each cluster, is the most lenient cutoff at which we would still feel comfortable diagnosing PTSD. Most clients in our clinic score well above this minimum threshold and clearly meet diagnostic criteria, but if someone scored below this level we would also feel confident in stating that they did not meet the DSM-IV diagnostic criteria for PTSD.
3. A Mississippi Scale. On occasion, the time available for a consultation is quite limited, and the referral question may simply be whether a patient should be referred for a more extensive evaluation. In this case, the assessment might consist of asking a patient to complete a PCL with respect to an index trauma, then briefly reviewing their answers with them.

Scenario 2: Comprehensive Clinical Assessment

Comprehensive assessments involve a more extensive evaluation of the areas noted above, with the goal of not only reaching a diagnosis of PTSD but of developing a rich case conceptualization. Additional components include:

1. A more detailed social history, with a more thorough exploration of traumatic events and other stressors across the life span. This is accomplished with a semi-structured clinical interview, along with the Evaluation of Lifetime Stressors (ELS; Krinsley, 1996), which consists of a screening questionnaire and a follow-up structured interview.
2. An MMPI-2, which provides information about PTSD (e.g., PK scale, F-2-8 profile) and other psychopathology, and which, most importantly, is the only instrument in our clinical battery with scales designed to formally assess response style (e.g., symptom exaggeration or minimization).
3. A more comprehensive assessment of comorbid psychopathology, by means of questionnaires and structured interviews. In terms of interviews, this component involves a SCID to assess mood, anxiety, psychotic, and substance use disorders, and, if warranted, a SCID-II to assess personality disorders. Questionnaires include the Beck Depression Inventory (BDI), the Beck Anxiety Inventory (BAI; Beck, Epstein, Brown, & Steer, 1988), the State-Trait Anger Expression Inventory (STAXI; Spielberger, Reheiser, & Sydeman, 1995), and the Symptom Checklist-90-Revised (SCL-90-R).
4. A psychophysiological evaluation, using either standard trauma-related stimuli, or standard stimuli plus idiographic stimuli based on the patient's unique trauma history.

Scenario 3: Survey with a Self-Report Instrument

In survey research the emphasis is typically on determining PTSD caseness in order to estimate prevalence rates and identify risk factors. Ideally, survey designs involve two stages: the survey stage, in which the survey instruments are administered to the full sample, and the validation stage, wherein a subset of respondents are also administered a diagnostic interview in order to calibrate the survey instruments for the sample. If the second stage is not feasible, it may still be possible to estimate prevalence in the survey sample by adopting scoring algorithms from a similar sample for which solid validity data are available. If neither of these approaches is possible, a rationally derived scoring algorithm may be used to estimate prevalence (e.g., scores of 2 or greater on a 5-point scale count as symptoms toward a diagnosis). However, this is much less satisfactory since it is not empirically justified for the specific sample.

For survey instruments consisting only of self-report measures, a thorny issue is how best to assess trauma. As discussed earlier, the questions are how to determine without an interview whether a respondent has experienced an event that would satisfy Criterion A and how to identify an event or events to serve as the focus for PTSD symptom inquiry. This is not a problem when a survey addresses the impact of a specific trauma or type of trauma, such as combat, rape, or disaster. But when a survey is designed to assess PTSD arising from any type of traumatic event, and respondents endorse multiple events on a self-report trauma checklist, they must be instructed as to which event to keep in mind as they rate their PTSD symptoms. Some possible approaches are:

1. Use PTSD measures that refer to a nonspecific traumatic event and don't try to link PTSD symptoms to a specific event, even if the survey includes a self-report measure of traumatic events. We would use the PCL, since it directly corresponds to the DSM-IV diagnostic criteria for PTSD and yields both a diagnosis and a dimensional severity score, and the Mississippi Scale, since it assesses associated features and has consistently emerged as the most valid self-report measure of PTSD. Depending on the target population, we would use either the civilian or military versions of these scales, and if feasible would calibrate them for the target population by administering a CAPS to a subsample of respondents.
2. Focus on a specific trauma population, such as combat veterans, rape victims, or disaster survivors, and instruct respondents to respond to items assessing PTSD symptoms with this index event in mind. In this case, we would use the specific version of the PCL

and ask respondents to write in the event they experienced. Again, depending on the target population, we would also use either the civilian or military version of the Mississippi Scale, and would calibrate it if possible against the CAPS.

3. Use a self-report trauma checklist, instructing respondents to choose the event they consider the worst and respond to items assessing PTSD symptoms with that event in mind. To assess trauma exposure, we would use an extended version of the LEC that assesses both parts of the DSM-IV definition of Criterion A, as well as the chronology, frequency, and type of exposure, for a broad range of traumatic events. For a study of war-zone trauma, we would also use the Combat Exposure Scale (CES). To assess PTSD symptoms, we would use the specific version of the PCL and the civilian version of the Mississippi Scale. The recently published PDS also appears promising for this application, as it is the only self-report PTSD measure that assesses all DSM-IV diagnostic criteria for PTSD, including Criterion A.

Scenario 4: Survey with an Interview-Based Instrument

Most of the considerations described in the previous scenario also apply when a survey is administered by an interviewer. The key differences are that the interviewer can (a) judge whether stressful events that the respondent describes meet Criterion A, (b) help the respondent select one or more events for symptom inquiry, and (c) determine if reported symptoms are attributable to a specific event.

In survey research the time allotted for an assessment is usually quite limited. Also, due to the expense of interviewing large samples, lay interviewers are typically used instead of trained clinicians. For these reasons, measures such as the DIS were designed for use in large-scale epidemiological investigations. The DIS is highly structured, with little if any follow-up or clarification permitted. As noted earlier, the DIS has been used in a number of survey studies of PTSD, and has proven useful for this application. However, one of its limitations is that symptoms are rated only as present or absent. We see no reason why, for at least some applications, lay interviewers couldn't be trained to administer interviews such as the PSS-I instead, which would yield both dichotomous ratings for symptoms and the PTSD diagnosis as well as dimensional scores reflecting PTSD symptom severity.

Scenario 5: Correlational Study

In some sense, any study involving PTSD and at least one other construct could be considered to be a correlational design. What we are referring to here, though, are studies comparing one or more measures of PTSD to measures of other constructs in an explicit correlational framework such as a multitrait-multimethod correlation matrix, a multiple regression equation, or a structural equation model. Our recommendations for this type of study are to use a structured interview for PTSD that yields dimensional severity scores for each symptom, and to use multiple measures of PTSD and all other constructs involved in the study. Self-report PTSD measures might include a DSM-correspondent scale and one or more PTSD-focused but non-DSM-correspondent scales. As a minimum battery for PTSD we would use the CAPS, the appropriate PCL, and the combat or civilian version of the Mississippi Scale. For large-scale assessment studies we would add the MMPI-2, since it would provide the PK scale as well as measures of other constructs and measures of response bias. We might also add the TSI, since it assesses a broad range of trauma-related outcomes and provides additional information about response bias.

Scenario 6: Case-Control Study

In case-control research the investigator assembles relatively homogeneous groups of participants and compares their characteristics or performance on some experimental task. Critical to this approach is the specification of explicit inclusion and exclusion criteria for the groups to be compared. In PTSD research, the PTSD group consists of participants who, at a minimum, unambiguously meet all of the DSM-IV diagnostic criteria for the disorder. There is also at least one control or comparison group composed of participants who unambiguously do *not* meet diagnostic criteria for PTSD. Other specifications might

involve type of trauma (e.g., rape victims, any childhood physical or sexual abuse, any Criterion A event), comorbid disorders (e.g., no current Axis I or II disorders, no current or lifetime psychotic disorder), demographic characteristics (e.g., females between ages 18 and 45), or abilities (e.g., reads at an eighth-grade level, not color-blind). Final decisions regarding inclusion and exclusion criteria are influenced not only by conceptual considerations, but also by the resources available for conducting the research, including the number of potential participants, the amount of clinician time involved, and the overall cost of the project. If there are restrictions on any of these resources, the criteria may need to be relaxed in order to obtain a sufficiently large sample.

In case-control research it is essential to conduct a structured diagnostic interview for PTSD. In addition, we strongly recommend administering a structured interview for other lifetime and current Axis I, and possibly Axis II, disorders. PTSD interviews that yield continuous severity scores are especially useful, since the threshold for inclusion can easily be adjusted based on the needs of the study, creating a more stringent or liberal threshold for a PTSD diagnosis. Including additional PTSD measures in the assessment battery allows the creation of very "clean" groups, by requiring that participants in the PTSD group exceed recommended cutoffs on all measures, while controls score below all cutoffs. This approach can be costly, however, as it will almost certainly render some potential participants ineligible for the study due to discordant PTSD indicators.

A recent study in our lab involving a simple two-group design illustrates some of the key considerations. All participants in the study were male Vietnam theater veterans with no history of psychotic disorders, all of whom were able to read the stimulus materials and had a mental status sufficient to complete an extensive assessment and experimental protocol. Participants in the PTSD group met diagnostic criteria for current PTSD on the CAPS according to the most stringent scoring rule, whereas controls fell below the diagnostic threshold according to the original, more liberal scoring rule. In addition, PTSD participants exceeded recommended cutoff scores on the PCL and the Mississippi Scale, whereas controls were below cutoffs on these measures. Finally, controls had no lifetime history of PTSD, either as a result of combat or a civilian trauma.

In terms of comorbidity, PTSD participants were included regardless of additional diagnoses (except for psychotic disorders), since we felt that a group of veterans with PTSD but no comorbid disorders would be unrepresentative. Controls were excluded if they had any current Axis I diagnosis, but no restrictions other than psychosis were put on lifetime disorders. To reduce the number of potential participants that were excluded from the study, a member of the research team conducted phone screens to determine probable eligibility. As a result of these inclusion and exclusion criteria, we created two groups that were quite similar in terms of demographic variables and level of combat exposure, but distinctly different in terms of PTSD symptoms.

Scenario 7: Treatment Study

In a treatment study, most of the diagnostic considerations discussed in the previous scenarios are applicable, including the need for a structured diagnostic interview, the need for explicit inclusion and exclusion criteria, and the use of multiple measures. An additional consideration is that multiple assessments are typically conducted, often over time intervals as brief as one or two weeks. This calls for a measure that can be administered quickly with minimal burden to clinician and respondent and that reflects the appropriate time frame (e.g., past week, past two weeks). Self-report measures are ideal for this purpose, especially DSM-correspondent measures, although some would require modification of the time frame to suit the specific assessment need. Interviews could be used, but are obviously more time-consuming and costly.

We would use a CAPS-DX to establish a PTSD diagnosis, using the original scoring rule. Although this rule may be too liberal for case-control research, it is still indicative of clinically significant PTSD symptomatology, and therefore is sufficient for inclusion in a treatment protocol. For brief assessment intervals over the course of treatment, we would use the CAPS-SX as the follow-up instrument, since it was

designed to assess PTSD symptom severity over the past week. Alternatively, if multiple interviews were not feasible for some reason, we might administer the CAPS-DX initially, then use the PCL over the course of treatment, conducting another CAPS-DX at the end of the study and at follow-up. Additional PTSD measures could also be administered at baseline and at follow-up.

SUMMARY AND CONCLUSIONS

As we hope this chapter has made clear, considerable progress has been achieved in the development and psychometric evaluation of standardized measures to assess trauma and PTSD. Clinicians and investigators can now choose from a broad array of PTSD measures, including self-report measures, structured interviews, and psychophysiological protocols. However, there is still room for improvement. One step that can be taken immediately is to adopt standardized, well-validated measures in every setting in which the assessment of traumatized individuals is conducted. There is no longer a compelling rationale for the use of nonstandardized approaches that do not yield quantifiable indicators of symptom severity. Nor should investigators create and use new measures of trauma and PTSD unless they identify a specific assessment need that is not addressed by an existing measure.

A second step, which will take many years to fully accomplish, is to conduct additional psychometric investigations of existing instruments to enhance the precision of measurement and generalization of findings across different settings and populations. More research on individual PTSD measures is needed to: (a) replicate and extend previous psychometric findings; (b) establish norms on a variety of populations; (c) identify the possible influence of gender, ethnicity, and cross-cultural differences on test scores; and (d) develop optimal scoring rules for addressing different assessment questions. Such research is particularly important to conduct when instruments are translated into languages other than the one in which they were developed. Investigators should not simply assume that translations are psychometrically equivalent to the original instruments, but should document this empirically. Studies are also needed in which multiple measures are administered to the same sample to identify the best measure for diagnosing PTSD in a given context and to determine how best to combine different measures into an optimal PTSD assessment battery.

These steps are essential for identifying the most reliable and valid measures and for enhancing the uniformity of PTSD assessment, but improved assessment is only a means to an end. The ultimate goals are to understand how traumatic events affect those who experience them and to develop better methods of treating PTSD and other posttraumatic sequelae. Advances in PTSD assessment over the past ten years have greatly facilitated progress toward these goals, and we are confident that the increasingly rich and mature assessment literature will continue to make a significant contribution to the scientific study of traumatic stress.

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). 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., Ward, C. H., Mendelson, M., Mock, J. E., & Erbaugh, J. K. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

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

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

Blanchard, E. B., Hickling, E. J., Taylor, A. E., Forneris, C. A., Loos, W., & Jaccard, J. (1995). Effects of varying

scoring rules of the Clinician-Administered PTSD Scale (CAPS) for the diagnosis of post-traumatic stress disorder in motor vehicle accident victims. *Behavior Research Therapy, 33,* 471–475.

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy, 34,* 669–673.

Breslau, N., Davis, G. C., Andreski, P., & Peterson, E. (1991). Traumatic events and posttraumatic stress disorder in an urban population of young adults. *Archives of General Psychiatry, 48,* 216–222.

Brett, B. (1993). Psychoanalytic contributions to a theory of traumatic stress. In J. P. Wilson & B. Raphael (Eds.), *International handbook of traumatic stress syndromes.* New York: Plenum Press.

Briere, J. (1995). *Trauma Symptom Inventory (TSI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cannon, D. S., Bell, W. E., Andrews, R. H., & Finkelstein, A. S. (1987). Correspondence between MMPI PTSD measures and clinical diagnosis. *Journal of Personality Assessment, 51,* 517–521.

Centers for Disease Control. (1988). Health status of Vietnam veterans: I. Psychosocial characteristics. *Journal of the American Medical Association, 259,* 2701–2707.

Cordova, M. J., Andrykowski, M. A., Kenady, D. E., McGrath, P. C., Sloan, D. A., & Redd, W. H. (1995). Frequency and correlates of posttraumatic-stress-disorder-like symptoms after treatment for breast cancer. *Journal of Consulting and Clinical Psychology, 63,* 981–986.

Davidson, J. (1996). *Davidson Trauma Scale* [Manual]. Ontario, Canada: Multi-Health Systems.

Davidson, J. R., & Foa, E. B. (1991). Diagnostic issues in posttraumatic stress disorder: Considerations for the DSM-IV. *Journal of Abnormal Psychology, 100,* 346–355.

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

Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring, & procedures manual-II.* Towson, MD: Clinical Psychometric Research.

Falsetti, S. A., Resnick. H. S., Resick, P. A., & Kilpatrick, D. G. (1993). The Modified PTSD Symptom Scale: A brief self-report measure of posttraumatic stress disorder. *Behavioral Assessment Review,* 161–162.

Foa, E. B. (1995). *Posttraumatic Stress Diagnostic Scale* [Manual]. Minneapolis, MN: National Computer Systems.

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.

Green, B. L. (1991). Evaluating the effects of disasters. *Psychological Assessment, 3,* 538–546.

Green, B. L. (1993). Identifying survivors at risk: Trauma and stressors across events. In J. P. Wilson & B. Raphael (Eds.), *International handbook of traumatic stress syndromes* (pp. 135–144). New York: Plenum Press.

Green, B. L., Epstein, S. A., Krupnick, J. L., & Rowland, J. H. (1997). Trauma and medical illness: Assessing trauma-related disorders in medical settings. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD.* New York: Guilford Press.

Green, B. L., Wilson, J. P., & Lindy, J. D. (1985). Conceptualizing PTSD: A psychosocial framework. In C. R. Figley (Ed.), *Trauma and its wake.* New York: Brunner/Mazel.

Hammarberg, M. (1992). Penn Inventory for posttraumatic stress disorder: Psychometric properties. *Psychological Assessment, 4,* 67–76.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7,* 238–247.

Helzer, J. E., Robins, L. N., & McEvoy, M. A. (1987). Post-traumatic stress disorder in the general population. *New England Journal of Medicine, 317,* 956–960.

Herman, D. S., Weathers, F. W., Litz, B. T., & Keane, T. M. (1996). Psychometric properties of the embedded and stand-alone versions of the MMPI-2 Keane PTSD Scale. *Assessment, 3,* 437–442.

Herman, J. L. (1992). *Trauma and recovery.* New York: Basic Books.

Horowitz, M. J. (1976). *Stress response syndromes.* New York: Jason Aronson.

Horowitz, M. J., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41,* 209–218.

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., Fairbank, J. A., Caddell, J. M., Zimering, R. T., Taylor, K. L., & Mora, C. A. (1989). Clinical evaluation of a measure to assess combat exposure. *Psychological Assessment, 1,* 53–55.

Keane, T. M., & Kaloupek, D. G. (1997). Comorbid psychiatric disorders in PTSD: Implications for research. In R. Yehuda & A. McFarlane (Eds.), *Psychobiology of posttraumatic stress disorder.* New York: Annals of the New York Academy of Science.

Keane, T. M., Kolb, L. C., Kaloupek, D. G., Orrm S. P., Blanchard, E. B., Thomas, R. G., Hsieh, F. Y., & Lavori, P. W. (1997). *Psychophysiological measurement in the diagnosis of post-traumatic stress disorder: Results from a Department of Veterans Affairs cooperative study.* Manuscript submitted for publication.

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.

Keane, T. M., Wolfe, J., & Taylor, K. L. (1987). Posttraumatic stress disorder: Evidence for diagnostic validity and methods of psychological assessment. *Journal of Clinical Psychology, 43,* 32–43.

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry, 52,* 1048–1060.

Kilpatrick, D. G. (1988). Rape Aftermath Symptom Test. In M. Hersen & A. S. Bellack (Eds.), *Dictionary of behavioral assessment techniques.* Oxford: Pergamon Press.

Kilpatrick, D. G., & Resnick, H. S. (1993). PTSD associated with exposure to criminal victimization in clinical and community populations. In J. R. T. Davidson & E. B. Foa (Eds.), *PSTD in review: Recent research and future directions.* Washington, DC: American Psychiatric Press.

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, 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 Publications.

Krinsley, K. E. (1996). Psychometric review of The Evaluation of Lifetime Stressors (ELS) Questionnaire & Interview. In B. H. Stamm (Ed.), *Measurement of stress, trauma, and adaptation.* Lutherville, MD: Sidran Press.

Krinsley, K. E., & Weathers, F. W. (1995). The assessment of trauma in adults. *PSTD Research Quarterly, 6,* 1–3.

Kulka, R. A., Schlenger, W. E., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. R., & Weiss, D. S. (1990). *Trauma and the Vietnam War generation: Report on the findings from the National Vietnam Veterans Readjustment Study.* New York: Brunner/Mazel.

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.

Lauterbach, D., & Vrana, S. (1996). Three studies on the reliability and validity of a self-report measure of posttraumatic stress disorder. *Assessment, 3,* 17–25.

Lerner, F. (1994). *PILOTS database: User's guide* (2nd ed.). White River Junction, VT: National Center for Post-Traumatic Stress Disorder.

Levin, P., & Reis, B. (1997). Use of the Rorschach in assessing trauma. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD.* New York: Guilford Press.

Litz, B. T., & Weathers, F. W. (1994). The diagnosis and assessment of post-traumatic stress disorder in adults. In M. B. Williams & J. F. Sommer, Jr. (Eds.), *Handbook of post-traumatic therapy.* Westport, CT: Greenwood Press.

Lyons, J. A., & Keane, T. M. (1992). Keane PTSD Scale: MMPI and MMPI-2 update. *Journal of Traumatic Stress, 5,* 111–117.

Lyons, J. A., & Scotti, J. R. (1994). Comparability of two administration formats of the Keane Posttraumatic Stress Disorder Scale. *Psychological Assessment, 6,* 209–211.

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.

Norris, F. (1992). Epidemiology of trauma: Frequency and impact of different potentially traumatic events on different demographic groups. *Journal of Consulting and Clinical Psychology, 60,* 409–418.

Resnick, H. S., Kilpatrick, D. G., Dansky, B. S., Saunders, B. E., & Best, C. L. (1993). Prevalence of civilian trauma and posttraumatic stress disorder in a representative national sample of women. *Journal of Consulting and Clinical Psychology, 61,* 984–991.

Resnick, H. S., Kilpatrick, D. G., & Lipovsky, J. A. (1991). Assessment of rape-related posttraumatic stress disorder: Stressor and symptom dimensions. *Journal of Consulting and Clinical Psychology, 3,* 561–572.

Resnick, P. H. (1997). Malingering of posttraumatic disorders. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS) and professional manual.* Odessa, FL: Psychological Assessment Resources.

Saigh, P. A., & Bremner, J. D. (1998). The history of posttraumatic stress disorder. In P. A. Saigh & D. J. Bremner (Eds.). *Posttraumatic stress disorder: A comprehensive textbook.*. Boston, MA: Allyn & Bacon.

Saunders, B. E., Arata, C. M., & Kilpatrick, D. G. (1990). Development of a crime-related posttraumatic stress disorder scale for women within the Symptom Checklist-90-Revised. *Journal of Traumatic Stress, 3,* 439–448.

Schlenger, W. E., Kulka, R. A., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. R., & Weiss, D. S. (1992). The prevalence of post-traumatic stress disorder in the Vietnam generation: A multimethod, multisource assessment of psychiatric disorder. *Journal of Traumatic Stress, 5,* 333–363.

Schwarzwald, J., Solomon, Z., Weisenberg, M., & Mikulincer, M. (1987). Validation of the Impact of Event Scale for psychological sequelae of combat. *Journal of Consulting and Clinical Psychology, 55,* 251–256.

Shear, M. K., & Maser, J. D. (1994). Standardized assessment for panic disorder research. *Archives of General Psychiatry, 51,* 346–354.

Solomon, S. D., & Canino, G. J. (1990). Appropriate of DSM-III-R criteria for posttraumatic stress di der. *Comprehensive Psychiatry, 31,* 1–11.

Spielberger, C. D., Gorsuch, R. L., & Luschene, E. (19 *Manual for the State-Trait Anxiety Interview (evaluation questionnaire).* Palo Alto, CA: Consul Psychologists Press.

Spielberger, C. D., Reheiser, E. C., & Sydeman, (1995). Measuring the experience, expression, control of anger. In H. Kassinove (Ed.) *Anger di ders: Definition, diagnosis, and treatment.* Wash ton, DC: Taylor & Francis.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & F M. B. (1990). *Structured Clinical Interview for D III-R.* Washington, DC: American Psychiatric Pre

Stamm, B. H. (Ed.). (1996). *Measurement of stress, trai and adaptation.* Lutherville, MD: Sidran Press.

Vrana, S., & Lauterbach, D. (1994). Prevalence of matic events and post-traumatic psychological sy toms in a nonclinical sample of college stud *Journal of Traumatic Stress, 7,* 289–302.

Vreven, D. L., Gudanowski, D. M., King, L. A., & K D. W. (1995). The civilian version of the Missis PTSD Scale: A psychometric evaluation. *Journ Traumatic Stress, 8,* 91–109.

Watson, C. G., Juba, M. P., Manifold, V., Kucala, T Anderson, P. E. D. (1991). The PTSD Interview: tionale, description, reliability, and concurrent v ity of a DSM-III-based technique. *Journal of Cli Psychology, 47,* 179–188.

Watson, C. G., Kucala, T., & Manifold, V. (1986). A c validation of the Keane and Penk MMPI Scal measures of post-traumatic stress disorder. *Journ Clinical Psychology, 42,* 727–732.

Weathers, F. W., Blake, D. D., Krinsley, K. E., Haddad Huska, J. A., & Keane, T. M. (1997). *Reliability validity of the Clinician-Administered PTSD S* Manuscript submitted for publication.

Weathers, F. W., Keane, T. M., King, L. A., & King, I (1997). Psychometric theory in the developme posttraumatic stress disorder assessment tools. In Wilson & T. M. Keane (Eds.), *Assessing psycho cal trauma and PTSD.* New York: Guilford Pres

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. Keane, T. M. (1993, October). *The PTSD Che (PCL): Reliability, validity, and diagnostic utilit* per presented at the annual meeting of the Int tional Society for Traumatic Stress Studies, Antonio, CA.

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.

Weathers, F. W., Litz, B. T., Herman, D. S., King, D. W., King, L. A., Keane, T. M., & Huska, J. A. (1997). *Three studies on the psychometric properties of the PTSD Checklist (PCL).* Manuscript submitted for publication.

Weathers, F. W., Meron, A., & Keane, T. M. (1997). *Psychometric properties of five scoring rules for the Clinician-Administered PTSD Scale (CAPS).* Poster submitted to the 13th Annual Meeting of the International Society for Traumatic Stress Studies.

Weiss, D. S. (1997). Structured clinical interview techniques. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD.* New York: Guilford Press.

Weiss, D. S., & Marmar, C. R. (1997). The Impact of Event Scale-Revised. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD.* New York: Guilford Press.

Wilson, J. P. (1995). The historical evolution of PTSD diagnostic criteria: From Freud to DSM-IV. In G. S. Everly, Jr. & J. M. Lating (Eds.), *Psychotraumatology.* New York: Plenum Press.

Wilson, J. P., & Keane, T. M. (Eds.). (1997). *Assessing psychological trauma and PTSD.* New York: The Guilford Press.

World Health Organization (1990). *International classification of diseases and related health problems* (10th ed.). Geneva, Switzerland: World Health Organization.

Zilberg, N. J., Weiss, D. S., & Horowitz, M. J. (1982). Impact of Event Scale: A cross-validation study and some empirical evidence supporting a conceptual model of stress response syndromes. *Journal of Consulting and Clinical Psychology, 50,* 407–414.