Ethnocultural Aspects of Posttraumatic Stress Disorder

Issues, Research, and Clinical Applications

EDITED BY

Anthony J. Marsella

Matthew J. Friedman

Ellen T. Gerrity

Raymond M. Scurfield

AMERICAN PSYCHOLOGICAL ASSOCIATION
WASHINGTON, DC

Copyright © 1996 by the American Psychological Association. All rights reserved.
Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

First printing February 1996
Second printing September 1996
Third printing August 1997

Published by
American Psychological Association
750 First Street, NE
Washington, DC 20002

Copies may be ordered from
American Psychological Association
Order Department
P.O. Box 92984
Washington, DC 20090-2984

In the UK and Europe, copies may be ordered from
American Psychological Association
3 Henrietta Street
Covent Garden, London
WC2E 8LU England

Typeset in Goudy by PRO-IMAGE Corporation, Techna-Type Div., York, PA

Printer: Port City Press, Baltimore, MD
Cover and Jacket Designer: Minker Design, Bethesda, MD
Jacket Illustrator: Stephanie Shieldhouse, Jacksonville, FL
Technical/Production Editor: Sarah J. Trembath

Library of Congress Cataloging-in-Publication Data
Ethnocultural aspects of posttraumatic stress disorder : issues, research, and clinical applications / edited by Anthony J. Marsella . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 1-55798-319-4 (acid-free paper)
1. Post-traumatic stress disorder—Cross-cultural studies.
2 Psychiatry, Transcultural. I. Marsella, Anthony J.
RC552.P67E83 1996
616.85'21—dc20

95-40252
CIP

British Library Cataloguing-in-Publication Data
A CIP record is available from the British Library.

Printed in the United States of America

7

ETHNOCULTURAL CONSIDERATIONS IN THE ASSESSMENT OF PTSD

TERENCE M. KEANE, DANNY G. KALOUPEK, and FRANK W. WEATHERS

Assessing the occurrence of potential traumatic events and their psychological sequelae, including posttraumatic stress disorder (PTSD), has become an important component of clinical care and research in the field of mental health (Wilson & Raphael, 1993). Since 1980, when PTSD was first included in the diagnostic nomenclature of the American Psychiatric Association (i.e., *DSM–III*; American Psychiatric Association, 1980), the study of this disorder has resulted in a more comprehensive understanding of the effects of major life stressors on adjustment, adaptation, and social functioning.

There also has been a parallel growth in the appreciation of the role of traumatic events in the development of other forms of psychopathology (e.g., substance abuse, personality disorders, and depression). Traumatic events may have a direct etiological link to psychopathology, may serve to precipitate certain problems, or may exacerbate already existing conditions. In addition, as Keane and Wolfe (1990) have pointed out, the development of PTSD can result in a cascade of psychological problems that reach criteria for other diagnostic conditions (e.g., substance abuse, major depressive disorder, panic disorder). Much of the progress and increased understanding of PTSD and its implications for psychopathology can be traced to the development of reliable and valid methods for assessing potential traumatic events and PTSD itself (Keane, Weathers, & Kaloupek, 1992).

The psychological assessment of PTSD and related factors has developed primarily within the context of Western, developed, industrialized countries (DeGirolamo, 1992; see also chapter 4, this volume). The purpose of this chapter is to identify issues related to ethnocultural considerations in the assessment of PTSD. To accomplish this objective we will examine existing models for conceptualizing the PTSD assessment process. Next, we will review fundamental procedures that are used in documenting the psychometric properties of measures of PTSD and then selectively review the literature on PTSD instruments that have taken ethnic and minority issues into consideration during the process of development.

We intend to demonstrate the process that might be considered when undertaking research and clinical work across cultures or across ethnic groups within a society. Marsella, Friedman, and Spain (1992) and Wilson and Raphael (1993) have reported that cross-cultural issues in PTSD are receiving increased attention world wide because of the high rates of war, natural disaster, ethnic conflict, and technological disaster in contemporary society. There is a clear need to develop methods that are suitable for diverse contexts and populations. In this chapter we will attempt to provide a framework for that development.

MULTIDIMENSIONAL ASSESSMENT OF PTSD

Keane, Wolfe, and Taylor (1987) outlined a method for evaluating the psychological effects of trauma exposure that incorporated assessment of the stressor variable (e.g., rape, disaster, combat), broad-based symptom measurement, and the use of multiple indices of PTSD. This methodological approach to assessment included recommendations to obtain information from the affected individual, collateral informants, available records, psychological tests, and psychophysiological indices. While a carefully conducted and sensitive clinical interview was the cornerstone of this approach to assessment, the addition of multiple indicators of symptomatology and adjustment was conceptualized as providing a more comprehensive understanding of the traumatized individuals and the contexts in which they functioned.

Recognizing the fallibility of any single indicator of PTSD and the likelihood that not all indicators would agree, we suggested the use of clinical judgment to reconcile any discrepancies found in the concordance of measures. Under research conditions, clinical judgments could be replaced by statistical algorithms, as was done in the National Vietnam Veterans Readjustment Study (Kulka et al., 1990). However, the benefit of a statistical versus a clinical decision-making process has been debated in multiple arenas.

With respect to targeting goals and objectives for treatment, Keane et al. (1987) recommend the use of clinical input by experienced clinicians to reconcile differences among the indicators and to arrive at a decision regarding the presence, absence, and severity of PTSD symptomatology. With the practice of consensus diagnosis, a patient can be classified optimally according to diagnostic criteria, and through this process of diagnosis, a functional analysis of the individual's problems can occur with suggestions for clinical interventions.

Research on the assessment of trauma exposure and PTSD has progressed remarkably in the past 15 years (Sutker, Uddo-Crane, & Allain, 1991). Clinical interviews for examining the diagnostic criteria have been developed and validated; psychometric measures have been constructed and tested empirically; and measures of psychophysiology have become increasingly accepted as a component of the comprehensive assessment of PTSD (Newman, Kaloupek, & Keane, in press). Because little evidence currently exists regarding ethnocultural measures of trauma exposure and PTSD, we will provide a brief overview of the available literature that has addressed the topic of cross-cultural assessment and trauma work.

There are some notable precedents that have been established for guiding the development of culturally sensitive instruments for trauma exposure and PTSD, and there are numerous ongoing research studies that have employed state-of-the-art methods in identifying and measuring cultural variations in the manifestation of the traumatic response. Our review will attempt to guide researchers and clinicians who wish to extend current knowledge in PTSD assessment to a specific cultural context. Our efforts will be directed toward elucidating scientifically sound methods of constructing measures to be used in PTSD work and the principles that would contribute to the development of sound, empirically based cross-cultural measures of this disorder.

CONSIDERATIONS IN DEVELOPING CROSS-CULTURALLY SENSITIVE PTSD MEASURES

In this section we will identify, define, and explain the many factors that researchers may wish to consider when developing measures of trauma and stress for use in research studies across cultures.

The Importance of Equivalency

To ensure that psychological assessment instruments are measuring the intended constructs across different cultures and societies, researchers and clinicians have employed numerous scientifically based methods, in-

cluding methods to assess conceptual equivalency (e.g., Dana, 1993; Marsella & Kameoka, 1988) and to assess reliability and validity. Applying these methods will increase the likelihood that a measure will provide the researcher or clinician with the desired information (e.g., the severity of PTSD or its prevalence). Ignoring the need for assessing the cultural relevance for a test or measurement instrument will be reflected in reliability and validity studies. If such studies are not conducted, the results of research may have adverse consequences for the individuals assessed, particularly if public policy is influenced by a study using faulty measures.

For example, a problem might arise if a PTSD measure that was developed in North America was adopted for use in an African culture and was simply insensitive to the cultural experience of trauma for the individuals. It is possible that the presentation of symptoms and the meaning and consequences of trauma could differ in the African culture. Thus, low rates of PTSD might emerge in a study and influence the researchers into erroneously concluding that rates of trauma were low, when in fact the measure itself might not assess the essence of traumatization for the culture under investigation. This hypothetical example amplifies the need for psychometric science in conducting cross-cultural research that has public policy implications.

There is a methodology available that, if applied with rigor, can avert situations such as this. Measures can be developed directly on the population to be studied (e.g., through the use of ethnoscience methods) or the performance of available instruments can be examined, either prior to their use or during the course of their use in research and clinical studies (e.g., pilot studies). In this way researchers and clinicians can be more confident that their instruments are performing in the intended manner, thus avoiding or reducing measurement error.

Kinds of Equivalence

Equivalence of instruments is a central tenet of cross-cultural assessment (Flaherty, et al., 1988; Kleinman & Good, 1985; Marsella, 1987; Marsella & Kameoka, 1988). It refers to the extent to which an assessment instrument fulfills its promise across different cultures and subcultures within a society. As specified by Marsella and Kameoka (1988), equivalence has many forms, and documenting equivalence increases the confidence that one has in the instrument to be used. The various forms of equivalence include (a) content equivalence, (b) linguistic and semantic equivalence, (c) conceptual equivalence, (d) scale and technical equivalence, and (e) normative equivalence.

Content Equivalence

Individuals in different cultures experience life and its vicissitudes in a variety of ways. If researchers were interested in response to rape across two distinct cultures, developing instruments for such a research project would involve a complex process because *rape* may be defined differently across cultures because of variations in gender relationships, sexual behavior, and perceptions of violence.

To construct an instrument for measuring PTSD within a culture, one needs to consider the various ways in which an individual in that culture might perceive, evaluate, and experience a high-magnitude stressor. To assist in this process, one might consider the creation of focus groups, a common contemporary method for assisting behavioral scientists in understanding a social problem and its implications for targeted populations.

A focus group for creating an instrument to understand the broad behavioral and psychological effects of rape in a society, for instance, might include people who have been raped, professional or respected leaders who have been involved in ministering to these people, and professionals familiar with the culture, language, and mores of the society. This group would attempt to generate a comprehensive list of the signs and symptoms associated with rape in that culture. Items from this list could be included in the development of an assessment instrument for measuring the effects of rape in that society.

If two societies vary in the way in which their citizens respond to rape, the items in the content domain will vary accordingly. It may prove difficult to devise a single instrument that would measure rape reactions fairly and equitably across both societies. If one were intent on developing a single instrument, the final product would at least need to reflect the specific reactions of the members of both cultures to properly measure and compare the adverse effects of rape for each group. The measure would include those items the cultures held in common as well as those items that were unique to one culture.

In theory, measures might need to be completely different to accomplish such goals across two cultures if the two cultures' responses to rape were completely different. However, if content equivalence is high for two cultures, then the same instrument would be acceptable. If content equivalence is low for two cultures, then separate instruments would be required.

Linguistic and Semantic Equivalence

If the same instrument is to be used in different cultures, as is often the case for expediency, researchers must establish semantic equivalence. This might involve translating an instrument from the parent language and then back-translating to see if the original meaning is retained. Translators

would work together to understand the concept under discussion and the various means of expressing the concept in the different languages. Grammar, denotative meanings, and connotative meanings should be similar or equivalent in the final product. One-way translation simply does not offer the opportunity to generate semantically or linguistically equivalent instruments that are capable of comparably measuring the construct under investigation in two cultures. The ultimate objective of semantic equivalence is to have the meaning of the statement in the test be the same across the two languages.

Conceptual Equivalence

Determining whether a concept is equivalent across different cultures is a difficult task. Marsella (1987) has described a series of steps that investigators can employ to increase the conceptual equivalence of their measurement tools across societies. These steps include, but are not necessarily limited to: (a) eliciting the domain of the concept (e.g., rape responses) through interviews with appropriate members of the society to be studied; (b) categorizing the items through an acceptable means of sorting, ranking, and scaling; (c) examining and rating the meaning of the items that have been generated through word association and antecedent–consequent methods; and (d) identifying the behavioral referents of the concept being examined.

Concepts such as anxiety, depression, aggression, anger, intrusive thoughts, and emotional numbing are central to understanding the psychological response to high-magnitude stressors. These characteristics have different meanings, and their presence has different implications for societies the world over. To assume cross-cultural conceptual equivalence of items measuring these constructs and symptoms surely would lead to mismeasurement of the important concepts related to traumatization in different societies. Even concepts that are related to traumatization—such as intimacy and love, suicide, death, and social support—may vary in different cultures so that variables that are known risk factors and correlates of traumatization in Western culture pose a major challenge for cross-cultural assessment.

Scale and Technical Equivalence

Scale and technical equivalence refers to the use of comparable metric methods in measuring a construct. Western cultures employ Likert scaling procedures and true/false methods from the time children are taught to read and write. As a result, there is a familiarity with this type of measurement of attitudes, behaviors, and experiences. This is less true for other societies. The types of options available for responding simply may not reflect the experience or the view of life of individuals from other cultures.

This problem is compounded by the predilection of the members of some cultures to respond in deference to the interviewer, rather than to avoid responding to items that seem unacceptable or that invalidly represent their personal experience.

Scales are not necessarily the same throughout the world; a seven-point Likert scale may function perfectly well in France and Sweden yet be incomprehensible in an African or Southeast Asian country. For this reason, efforts to ascertain the technical equivalence of a test will enhance the quality of the data obtained in a research or clinical setting.

Normative Equivalence

Norms are often the bases for determining abnormality and normality in specific cultures. Normative standards developed in one society may not necessarily generalize to other societies, however. Attempting to generalize norms across societies that differ in known ways on many characteristics can lead to inaccurate conclusions regarding the presence, absence, or severity of symptoms.

More to the point, making comparisons between non-Western individuals' performance on tests or instruments developed with Western norms risks two major problems. First, it gives precedence to the Western way of experiencing life. Second, it does not inform the investigator of the relative status of individuals compared to their ethnic and cultural peers.

Summary of Cross-Cultural Equivalence

The concept of cultural equivalence in personality testing is more a hope than a reality. Nonetheless, it remains an important objective for all researchers to attempt to reach equivalence on each of the dimensions discussed previously. The use of culturally fair tests is indicative of a thoughtful research program that is likely to yield valuable and maximally useful information. This is true whether the research is across Eastern and Western cultures or within a multiethnic country such as Australia, Canada, or the United States.

Recent modifications of the Minnesota Multiphasic Personality Inventory (MMPI-2; Hathaway & McKinley, 1989) have highlighted the importance of considering ethnicity, language, and experience in the revision of specific questions in the test. In addition, the restandardization samples also included normative groups from major ethnic subsections of society. This effort is a clear improvement over previous development methods used for the MMPI.

However, using a single variable such as ethnicity on which to draw conclusions regarding a group of people (e.g., African Americans, Hispanic Americans, Asian Americans) presents as many problems as it solves. The primary difficulty is that such clustering ignores individual differences

within such groupings in experience, constitution, and acculturation, while assuming homogeneity among large groups of people. Careful attention to the specific factors that underlie the ethnicity variable and contribute to observed differences between groups ultimately may prove to be a more fruitful approach to understanding individual differences in psychological status (e.g., discrimination, poverty, religious beliefs, educational sophistication, etc.).

STRATEGIES FOR CROSS-CULTURAL RESEARCH

Universality or Relativism in Psychopathology

There are two strategies recommended for making valid comparisons across cultures in the study of psychopathology. Neither strategy is entirely satisfactory, and the field is still struggling to address methodological issues and problems inherent in each. The demands of society and the need for public policy often drive cross-cultural research, although the field's research methods may not be up to the task yet.

One strategy has been to employ standard research instruments across cultures, translating them into the best format possible. The premise of this approach is that there are fundamental characteristics of mental disorders that transcend culture, race, and ethnicity. The instrument, if carefully translated for maximal equivalence, is assumed to be able to detect cases of the disorders targeted for study.

However, the assumption of universality of psychopathology has been shattered by numerous researchers over the years (Kleinman, 1977; Kleinman & Good, 1985; see also chapter 4, this volume). Perhaps this premise of universality is wishful thinking on the part of researchers. An alternative approach is to examine the statistical and psychometric properties of measures used across cultures to estimate the degree to which instruments are working in comparable ways across societies through techniques such as factor analysis (Kameoka, 1985; Marsella & Kameoka, 1988; Marsella, Kinzie, & Gordon, 1973).

For example, in a pioneering study of cross-cultural depression, Marsella, Kinzie, and Gordon (1973) submitted a depression checklist administered to Caucasian Americans, Chinese Americans, and Japanese Americans to a factor analysis according to each ethnic group and found variations in factor structure across the groups. The Chinese group showed a strong somatic factor; the Japanese a strong interpersonal factor; and the Caucasians a strong existential factor orientation. These were interpreted to reflect primary dimensions of self-construction and presentation. Although this is a reasonable first step, it is simply the first step in what is undoubtedly a complicated multiphasic process.

Looking at factor structures, interitem analyses, coefficient alphas, and so forth for an instrument employed across cultures will give a sense of the instrument's reliability, a fundamental basis from which one can begin to look at validity. However, the fact that an instrument is reliable and its items seem to interrelate in comparable ways across cultures does not ensure that it is measuring the same construct across groups of people. Validity studies also must be undertaken.

Validity studies may take place within the context of the main study, either preceding it as in a preliminary validation study or following the main study as in the hierarchical approach recommended by Dohrenwend and Shrout (1981). In the former strategy, specific steps are taken to determine the extent to which the instrument is measuring the construct of interest prior to the initiation of the main study. In the latter approach, clinical interviews by those knowledgeable in the cultural manifestations of the targeted disorder may be used to substantiate the presence, absence, and severity of the disorder and correlate these findings with those observed using the instrument employed in the main study.

The second approach employed in cross-cultural studies is the development of *emic* instruments (see Kleinman & Good, 1985). An emic instrument is one that is consistent with the indigenous views of psychopathology in the culture to be studied. This approach addresses the need for precision in measurement of a psychological construct within a culture but presents additional problems when cross-cultural comparisons are attempted. An item that is meaningful in one culture may not at all represent another culture's perspective on deviance or pathology.

Some have recommended that outcome measurement (i.e., disorder) should be related in systematic ways to other variables and that theoretical models should be able to predict poor psychological functioning regardless of the precise symptoms or characteristics measured. For example, a measure of the psychological effects of rape in two cultures actually may include different symptoms and outcomes, but these outcomes would be related in certain predictable ways to societal functioning, interpersonal functioning, and psychological functioning. Models could be created to determine if a measure of rape's effects in one culture is related to functioning in that society as much as a second measure of rape's effects relates in functioning in that society.

The latter approach ultimately may prove to be the most accurate in comparing diverse societies on symptoms of psychopathology. Unfortunately, it has the limitation of extreme statistical and theoretical complexity, a problem of serious importance when the people using the data are policy makers, politicians, and citizens rather than social or behavioral scientists. Often, such individuals prefer a simple answer to complicated questions to justify any actions taken on behalf of the denizens of a country or a subculture. Providing realistically complex answers may be discouraged

by those who are trying to make decisions about the actual distribution and use of resources.

Recommended Steps to Improve Ethnocultural Sensitivity

Marsella and colleagues (1985) summarized their thinking about the conduct of cross-cultural studies of psychopathology, recommending that scientists should (a) use relevant anthropological data to determine comparable life patterns and experiences in each culture, (b) develop glossaries and definitions of various deviant psychological experiences, (c) derive symptom patterns and interrelationships with statistical methods, (d) use similar sampling and case identification methods, (e) develop culturally based assessment instruments that optimize equivalence, and (f) establish baselines of normal functioning for the measures in different cultures.

These suggestions remain as pertinent today as they were a decade ago; yet there are few studies to date in the literature that employ these recommendations. Perhaps the cost associated with the conduct of such studies outweighs the benefits accrued. Perhaps the importance of cross-cultural research diminishes as society places greater emphasis on interventions for the individuals in need. Perhaps the demands of research within cultures exceeds the demand for comparisons across the cultures. Regardless of the problems, there will continue to be both a need and an interest in understanding how cultures experience psychopathology, how the members of a culture study it optimally within their culture, and how researchers can learn from other societies to minimize disorder and promote prosocial behavior on the part of individuals in all societies.

PSYCHOMETRIC CONSIDERATIONS IN ETHNOCULTURAL ASSESSMENT

Researchers and clinicians who embark on the study of psychopathology in general or PTSD in particular among populations that differ in cultural background must be prepared to apply the same standards to their work with these populations that they might with a majority population. Developing or using psychological tests or diagnostic instruments of any type requires an appreciation for the psychometric properties of an instrument. In this section, we present the notions of reliability and validity and how they should be used to evaluate the performance of a particular instrument in a specific population.

Validity

Validity is the extent to which an instrument is measuring what it purports to measure. To arrive at an estimate of validity one can employ

several different procedures, depending on interest, resources, and available collateral measures. Criterion validity is the relationship of the individual's performance on the instrument to some external referent of interest or behavior of interest. If the criterion is a clinician's judgment of the presence of a psychological disorder (e.g., PTSD, the correlation of the instrument with the judgment is a measure of criterion validity. If these measures are taken at the same time, this form of criterion validity is referred to as *concurrent validity*. If one measure is used to predict a future status or condition, this is referred to as a *measure of predictive validity*.

The key to criterion validity is the credibility and accuracy of the criterion. Using the quidelines stated previously, it is crucial that the criterion be one that is acceptable to the culture under study, preferably even gendered by members of the culture who have relevant status to the criterion under investigation (e.g., health care professionals, spiritual healers). A focus group, as discussed in an earlier section, is one method used to generate acceptable criterion standards. In the mental health field, items to be included in the criterion should reflect the known expressions of the disorder, the idioms of expression for that disorder, and any additional ethnocultural considerations in describing the disorder in question. The test or instrument under development should then correlate with the domain of symptoms and behaviors identified by the specialists who have established the criterion.

Content Validity

Content validity is the extent to which the measurement tool represents the total domain of the criterion. Since it is difficult to employ a test that contains all symptoms possible, most tests and instruments contain a subsample of the available possibilities. Content validity is a measure of the instrument's relationship to the entire domain of symptoms. Does a test or instrument adequately sample the universe of possible symptoms? An evaluation of the content validity of a test is made on the basis of expert judgment.

Construct Validity

Construct validity is employed when there is no universally acceptable criterion with which to compare the instrument or test being developed. The methods employed in construct validation are comparisons between the measure of interest with other existing measures of the construct under investigation. Does the new measure behave as one would predict it should? If so, then it may well be measuring the important construct. The more measures with which one can compare performance, the stronger the claims one can make for achieving a high degree of construct validity. In the area of psychopathology, and PTSD especially, construct validity is the

approach of choice when attempting to validate the performance of a measure to assess the psychological effects of exposure to massively stressful life experiences.

Construct validation has three distinct steps. First, the theoretical relationships among measures must be specified a priori. Second, the relationships among measures must be assessed properly using contemporary methods of statistical analysis. Third, the empirical relationships among measures must be interpreted in terms of clarifying the underlying construct of interest. These stages then can be repeated with different samples and under different conditions to increase the likelihood that the relationship observed in the earlier study is stable across time, samples, and conditions of administration.

Reliability

There are several methods of reliability (reproducibility) typically employed in studies of instrumentation development. Reliability sets an upper limit on validity. Thus, a thorough evaluation of reliability is a crucial step in the initial stages of developing a new instrument. Several different types of reliability might be assessed, each reflecting concerns about a different potential source of measurement error.

Test–Retest Reliability

Test–retest reliability addresses the replicability or stability of an instrument over time. It typically involves administering the same test to the same people at two different points in time and determining the relationship of performance at time 1 and time 2. A problem in using test–retest reliability in psychopathology research is that many disorders are cyclical and change over time. Thus, there is inherent change in symptom reporting, sometimes reflecting improvements in conditions and sometimes representing exacerbation of disorder. Selecting the appropriate interval for retesting becomes an important consideration. Most psychological disorders change gradually and, in the absence of treatment, an interval of a week or two typically would be optimal in assessing the stability of the measurement instrument over time.

Internal Consistency Reliability

Internal consistency methods are attempts to establish the replicability of scores obtained on multiple items in a test. Coefficient alpha is the most common form of internal consistency examined for psychological tests. It represents the relationship of scores on all items of a test to all

other items on the test to provide an estimate of how consistently the individuals in the sample responded to the questions.

When a test consists of a single construct, responses to the test items should be highly interrelated. If, however, a test measures many different constructs (or factors), coefficient alpha should be calculated separately on items reflecting a different factor or construct to determine the reliability of each independently. Alphas above .80 generally are considered acceptable estimates of the reliability of an instrument or its components.

Utility Analysis

Psychiatric conceptualizations of human behavioral problems historically have employed a medical model of disease. One implication of this is that individuals are classified dichotomously as either having or not having a condition. The *DMS–III* (American Psychiatric Association, 1980) attempted to incorporate continuous and qualitative measures into the diagnostic process through the use of the multiaxial system. However, the current zeitgeist for psychiatric disorders is to consider them as categories that require certain minimal-symptom expressions in order to exceed threshold. Thus, measures that yield dichotomous ratings of disorders (e.g., the presence or absence of PTSD) have been developed.

Diagnostic instruments and tests used for the express purpose of arriving at a clinical diagnosis also must be subjected to rigorous evaluation to determine their adequacy. Kraemer (1992) described in detail the various ways in which one can examine the extent to which a test does what it was intended to do by its developers. Her description of utility analyses is now the standard in the field, explaining methods for understanding the relationship of scores on a test to a standard for determining diagnostic category membership.

Sensitivity and Specificity

Sensitivity is one measure of diagnostic utility, and it is the extent to which a measure can identify true cases of a condition such as PTSD. *Specificity* is another measure of utility, and it represents the extent to which a test can identify true noncases. For instance, a test that identifies all known cases of PTSD in a sample would have optimal sensitivity. If that test correctly identified all individuals who did *not* have PTSD, it would have optimal specificity. Of course, in practice it is rare for any type of test to have perfect sensitivity and specificity, especially in mental health. However, the research for instruments that will provide excellent sensitivity and specificity for the purposes at hand (i.e., clinical work or

research) is a critical one in psychological work that extends across different cultures and ethnic groups.

Predictive Power

The predictive power of a positive test is also a measure of test utility, and it is defined as the true positives divided by the true positives plus false positives. In a similar way, the predictive power of a negative test is defined as true negative divided by true negatives plus false negatives. Efficiency is essentially a test's hit rate—that is, the proportion of cases that were correctly identified by the test. Each of the measures contained in a utility analysis provides incremental information about how a test performs.

Because there is a trade-off between sensitivity and specificity, the specific needs of an investigator or a clinician should determine which component of the utility analysis is more important given the particular assessment needs. If the goal is to maximize the detection of false negatives, a lower cutoff score usually would achieve this goal. This would lead to enhanced levels of sensitivity but lower specificity. If the goal is to minimize false positives, higher cutoffs would be used, resulting in greater specificity but lower sensitivity. Optimal sensitivity and specificity for tests generally is considered the goal for instruments (Kraemer, 1992), but these goals can vary depending on the needs of the investigator, the conditions of administration, and other variables impinging on the testing situation.

Summary of Psychometric Considerations

In constructing measures for culturally diverse populations, the primary considerations are the same as those for constructing any test with any population. Problems arise when researchers and clinicians attempt to overextend the known parameters of a test beyond the populations for which they were intended. Reliability and validity are primary concerns in understanding how an instrument will work with a specific population, and measures of reliability and validity will determine the quality of all work that will be accomplished subsequently with an instrument. In the case of psychiatric nosology, there is a definite need to incorporate comprehensive utility analyses so that the relative efficiency of a test can be determined on the samples or groups of individuals to be studied. To do less than this is to risk error in judgments about the presence, severity, and impact of psychological disorders in cultures of interest.

EXAMPLES OF RACIALLY AND CULTURALLY SENSITIVE ASSESSMENT INSTRUMENTS

There are may possible approaches that researchers and clinicians can take to increase the sensitivity of their assessment instruments to individ-

uals from other cultures or subcultures within a majority culture. Existing effective instruments can be properly translated, instruments can be modified to reflect cultural nuances for a disorder, or entirely new instruments can be developed. We will review examples of each of these approaches to addressing the problem of ethnocentricity in assessing symptomatology.

Indochinese PTSD Screening Scale

Mollica, Wyshak, de Marneffe, Khuon, and Lavelle (1987) present three versions of the Hopkins Symptom Checklist-25 (HSCL-25) that was translated concurrently into Khmer, Laotian, and Vietnamese. These researchers were looking for an effective screening tool for their clinic, which serves Indochinese refugees living in the United States who were exposed to the ravages of the many years of war in their home countries.

They employed translation and back-translation for the 25 items of the HSCL using individuals who were fluent in both English and their native languages and knowledgeable about mental health concepts. The next step was to administer the instrument to patients, gathering experience with the reactions of patients who were representative of the types of people for whom the test was intended. Then a group of cultural experts reviewed the instrument to ensure that the literal translations and connotations were identical in English and the Indochinese languages.

Mollica and colleagues (1987) found excellent test–retest reliability (r = .89) and interrater agreement data (r = .98). Validational studies demonstrated that the scores on the test were correlated with *DSM–III* diagnoses given by a clinician blind to test scores, and the authors also provided information on the sensitivity and specificity of the test (.88 and .73, respectively). The construction of these tests in three languages combined the best of many of the techniques discussed previously. Attending to the issues of culture, language, and psychometric rigor, these authors made a significant contribution to the literature on cross-cultural assessment. However, there were several problems. Small sample sizes raised questions about generalizability. Further, the researchers failed to assess the extent of the subjects' ethnic identification or attachment to their native cultures. Thus, it was possible that the subjects were not good representatives of their cultural traditions.

Harvard Trauma Scale

This same research group (Mollica et al., 1992) developed the Harvard Trauma Questionnaire, which combines the assessment of traumatic stressors (e.g., torture, rape, murder, starvation, etc.) and traumatic responses (e.g., anxiety, depression, nightmares, alienation, etc.) in one instrument. The questionnaire was developed in English by subject matter

experts and then it was translated into the three Indochinese languages of Khmer, Vietnamese, and Laotian by experienced mental health professionals. These professionals generated items for inclusion in the questionnaire based on their clinical experience with the refugees. The *DSM–III* criteria, supplemented by items from their knowledge of the refugee's condition, constituted the symptom items on the questionnaire. Unfortunately, little use was made of indigenous people in selecting the items. Thus, idiomatic expressions of problems were not included, resulting in possible false positives or negatives.

The scale was then back-translated to attain linguistic equivalence, and finally the entire group of experts engaged in a consensus discussion to create the final items for the questionnaire. Measures of test–retest reliability, interrater agreement, and internal consistency were computed following administration of the questionnaire to patients in a clinic. Test–retest reliability was .89 at 1 week; interrater agreement was .93; and coefficient alpha was .96 for the symptoms. The researchers then applied the HSCL-25 and the Harvard Team Questionnaire to their examination of the prevalence of mental health and health consequences for Indochinese refugees in a refugee camp in Thailand (Mollica et al., 1992). This sequence of studies demonstrated an approach to the development and use of both new instruments and modification of old assessment instruments for use across widely divergent cultures and populations.

Vietnamese Depression Scale

Kinzie and colleagues (1982) developed a depression measure for Vietnamese refugees initially using the Beck Depression Inventory (Beck, Ward, Mendelsen, Mock, & Erbough, 1961) and then adding cultural idioms and expressions for depressive mood. When applied to Vietnamese patients with known depression, the items from the inventory seemed less applicable to the patients than were the items related to cultural expression of depression. These researchers concluded that linguistic or semantic equivalence was inadequate in developing a measure of depression for this particular subculture within the United States. They included in their final instrument symptoms of depression that described affective, cognitive, and somatic components of depression within the context of the subjects' cultural experience of depression.

Measuring PTSD in African Americans

Penk and Allen (1991) described the relevant parameters for conducting a culturally sensitive assessment of African American veterans with PTSD. Their work represents a valuable description of the importance of distinguishing the effects of trauma on minority groups within Western

societies. Penk and Allen (1991) argued that minorities who are exposed to potentially traumatic events experience the events in different ways than do members of the majority culture. The result of this differing experience is the need for incorporating specific assessment methods when evaluating minorities for PTSD. Recognizing the role of prejudice, discrimination, and socioeconomic factors in assessing the effects of potentially traumatic events, according to the researchers' thesis, will result in a more clinically sophisticated understanding of the person presenting for treatment or research. They recommended that the use of instruments developed for use on the majority culture be standardized on the minority cultures prior to their use and interpretation.

When developing measures for use across subcultures within a majority culture (e.g., African Americans), there are several approaches that can be taken. An issue of continuing importance in these situations is the extent to which minorities are acculturated into the majority culture. However, measures of acculturation are few and most are controversial as to their validity.

However, preliminary reports from Marsella and his colleagues (Marsella, Horvath, & Tsushima, 1995) suggested that a behavioral index consisting of behaviors that subjects engage in that are related to traditional cultural practices seems to have good psychometric properties for distinguishing various levels of acculturation. A sound strategy to adopt in creating measures that are to be used across subcultures within a society is to include sufficient numbers of minority individuals in the validational studies conducted with the instrument.

Mississippi Combat-Related PTSD Scale

In constructing the Mississippi Scale for Combat-Related PTSD, Keane, Caddell, and Taylor (1988) oversampled African Americans for inclusion in the initial study of reliability. Of the 362 subjects in the first study, 45% were African American. Principal component analysis and calculation of coefficient alphas for the six factors extracted did not differ depending on the race of the subjects studied. Test–retest data were collected in a second study in which 18% of the total sample were minority veterans. In the third study of the validity of the scale, 40% of the PTSD group was African American. Inspection of the data on the Mississippi Scale indicated that there were no differences among the racial groups of subjects in the initial reliability and validity studies on this instrument. This may account for the success of this scale in both the clinical and epidemiological studies in which it has been used (Kulka et al., 1988; McFall, Smith, MacKay, & Tarver, 1990). While this scale appears to work well in the United States, its use with samples from other cultures and in different languages would necessitate additional validational data.

In a similar way, the Keane PTSD (PK) scale of the MMPI (Keane, Malloy, & Fairbank, 1984) was validated on more than 35% of African American subjects with PTSD in a total sample of 100. Their responses were compared to Caucasians with PTSD and to a sample of 100 comparison subjects with general psychiatric disorders. Inspection of the clinical and validity scales of the MMPI in this study revealed few differences between the groups on the scales in general and on the PK scale specifically (Fairbank, Caddell, & Keane, 1985). The use of the PK scale with this minority group seems empirically warranted by the date that have been collected. Its use with other minority groups and in studies in different cultures awaits additional confirmatory data.

Clinician-Administered PTSD Scale

The Clinician-Administered PTSD (CAPS; Blake et al., 1990; 1995) is a structured diagnostic interview for PTSD designed to be a comprehensive measure of PTSD. It assesses the 17 core symptoms of the disorder and 8 associated features, as well as the effects of specific symptoms on social and occupational functioning. A major strength of this scale is its behaviorally based anchors for all ratings and its guidelines for evaluating current and lifetime PTSD. In a series of studies to examine the reliability and validity of the CAPS conducted by Weathers and colleagues (1992), 25% of the participants were minorities.

In addition to demonstrating the excellent psychometric properties of this instrument, the authors also inspected racial differences in performance with the scale. There were no differences between Caucasian and African American subjects in total scores on the CAPS or on any subscale of the instrument, indicating the capacity of the instrument to assess PTSD across racial groupings. Construct validity studies indicated that the CAPS correlated .91 with the Mississippi Scale, .77 with the PK of the MMPI, and .89 with symptoms endorsed on the Structured Clinical Interview for *DSM–III* (SCID; Spitzer, Williams, Gibbon, & First, 1990). Test–retest reliability was .90; sensitivity rates were 84%; specificity was 95%; efficiency was 89%; and kappa against the SCID was .78.

PTSD Checklist

Using the same approach to evaluating the usefulness of an assessment instrument across ethnic groups, Weathers and colleagues (1993) developed the PTSD Checklist (PCL), a simple questionnaire that included the 17 *DSM–III* and *DSM–IV* symptoms associated with PTSD. Employing a five-point ratings scale, the PCL displayed excellent psychometric properties with coefficient alphas equal to .97, test–retest reliability at .96, and strong relationships with the Mississippi Scale (.93), the PK Scale (.77), and the

Impact of Event Scale (.90; Horowitz, Wilner, & Alvarez, 1979). Diagnostic utility measures of sensitivity (82%) and specificity (83%) both were strong.

An inspection of the performance of different racial groups on this self-report measure indicated that African Americans, who made up 23% of the sample, did not differ from Caucasians in their scores. Further, there were no statistically significant interactions between race and PTSD status. It seems that both the CAPS and the PCL are excellent measures of PTSD and that the measures function comparably across ethnic subcultures in the United States.

In the context of studying the effects of World War II on Dutch resistance fighters, Hovens (1994) and his colleagues conducted a sequence of studies to determine the reliability, validity, and utility of various measures of PTSD. Recognizing that instruments translated from English might not competently measure the essence of PTSD in Dutch, they employed a strategy whereby they both translated instruments using back-translation to promote semantic equivalence and then developed their own instruments in Dutch. The CAPS, the Mississippi Scale, the PK scale of the MMPI, and the SCID were all translated into Dutch and then empirically examined. Each of the instruments was subjected to a series of reliability, validity, and utility studies during the course of the researchers study of the resistance fighters. Their efforts to conserve resources by using extant measures, coupled with developing measures in their own language, is a model for cross-cultural instrument development.

CONCLUSION

The issues raised regarding cross-cultural assessment of PTSD are indeed daunting ones. To adhere to the recommended procedures is expensive, time consuming, and difficult work. Moreover, rarely is there financial support for instrument development and validation in the mental health field. Psychometric studies have little glory associated with them. However, they are the cornerstones for all work in psychopathology; they are *sine qua non*.

The model we have proposed for assessing PTSD (Keane et al., 1987), suggested that a single PTSD instrument should not be the basis for conferring a PTSD diagnosis, even if it is developed and evaluated carefully. All instruments are subject to measurement error, and the use of multiple indicators is recommended—particularly when important clinical or policy issues are involved. Employing a multimethod approach combining structured interviews, psychological tests, relevant archival data, and even psychophysiological assessment (cf. Prins, Kaloupek, & Keane, 1995) ultimately will provide the most reliable information possible. The

fundamental principle of using multiple measures for case identification is a necessity for cross-cultural research as well as research conducted within a single cultural context.

Further, simply translating instruments into the language of the targeted population is an inadequate strategy when conducting research or clinical work with diverse populations. Assessment of the translated instruments' psychometric properties is a minimal first step to determining the applicability of the instrument in the new language. Measures of reliability and validity must be collected to ensure that any data collected can be interpreted properly. When seeking a precise diagnosis for the purposes of a study, comprehensive utility analyses should be conducted to ascertain the extent to which the instruments employed are functioning in the desired manner.

There are no available assessment instruments for PTSD that can be used with impunity across all cultures and languages. There is much work to be done and much new knowledge needed before we can conduct successfully true cross-cultural studies. In the interim it would be wise for clinicians and researchers to appreciate the limits of our knowledge and to administer and interpret measures of personality and psychopathology in a cautious and culturally sensitive way. When opportunities present themselves to develop or modify instruments according to sound psychological principles, this should be the top research priority. Time and effort expended on instrumentation will yield huge benefits for the study at hand and for the entire field of cross-cultural psychology.

REFERENCES

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh J. (1961). An inventory of measuring depression. *Archives of General Psychiatry, 4*, 53–63.

Blake, D. D., Weathers, F. W., Nagy, L. N., Kaloupek, D. G., Gusman, F., 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. N., Kaloupek, D. G., Klauminser, G., Charney, D. S., & Keane, T. M. (1990). A clinician rating scale for assessing current and lifetime PTSD: the CAPS-1. *Behavior Therapist, 18*, 187–188.

Dana, R. (1993). *Multicultural assessment perspectives for professional psychology*. Boston: Allyn & Bacon.

de Girolamo, G. (1992). International perspectives on the treatment and prevention of posttraumatic stress disorder. In J. P. Wilson & B. Raphael (Eds.),

International handbook of traumatic stress syndromes (pp. 935–946). New York: Plenum Press.

Dohrenwend, B. P., & Shrout, P. E. (1981). Toward the development of a two-stage procedure for case identification and classification in psychiatric epidemiology. *Research in Community and Mental Health, 2*, 295–323.

Fairbank, J. A., Caddell, J. M., & Keane, T. M. (1985, August). *Black–White differences on the MMPI and PTSD subscale.* Paper presented at the Annual Convention of the American Psychological Association, Los Angeles.

Flaherty, J. A., Gaviria, F. M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J. A., & Birz, S. (1988). Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disease, 176*, 257–263.

Hathaway, S. R., & McKinley, J. C. Minnesota Multiphasic Personality Inventory — 2. Minneapolis: University of Minnesota.

Horowitz, M. J., Wilner, N. R., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective distress. *Psychosomatic Medicine, 41*, 208–218.

Hovens, J. E. (1994). *Research into the psychodiagnostics of posttraumatic stress disorder.* Amsterdam, Netherlands: Eburon Press.

Kameoka, V. A. (1985). Construct validation of psychological measures in cross-cultural research: Analysis of linear structural relationships. In R. Diaz-Guerrero (Ed.), *Crosscultural and national studies in social psychology* (pp. 88–99). Dordrecht, Netherlands: North Holland.

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., 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., Weathers, F. W., & Kaloupek, D. G. (1992). Psychological assessment of post-traumatic stress disorder. *PTSD Research Quarterly, 3*, 1–8.

Keane, T. M., & Wolfe, J. (1990). Comorbidity in post-traumatic stress disorder: An analysis of community and clinical studies. *Journal of Applied Social Psychology, 20*, 1776–1788.

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

Kinzie, J. D., Manson, S. M., Vinh, D. T., Tolan, N. T., Anh, B., & Pho, T. N. (1982). Development and validation of a Vietnamese-language depression rating scale. *American Journal of Psychiatry, 139*, 1276–1281.

Kleinman, A. (1977). Depression, somatization, and the "new" cross-cultural psychiatry. *Culture, Medicine, and Psychiatry, 6*, 1–39.

Kleinman, A., & Good, B. (Eds.). (1985). *Culture and depression.* Berkeley: University of California Press.

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

Kulka, R. A., Schlenger, W. E., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. R., & Weiss, D. S. (1988). *National Vietnam Veterans Readjustment Study (NVVRS): Description, current status, and initial PTSD prevalence estimates.* Research Triangle Park, NC: Research Triangle Institute.

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

Marsella, A. J. (1987). The measurement of depressive experience and disorders across cultures. In A. J. Marsella, R. Hirshifield, & M. Katz (Eds.), *The measurement of depression: Biological, psychological, behavioral, and social aspects* (pp. 376–397). New York: Guilford Press.

Marsella, A. J., Friedman, M. J., & Spain, E. H. (1992). A selective review of the literature on ethnocultural aspects of PTSD. *PTSD Research Quarterly, 3,* 1–8.

Marsella, A. J., Horvath, A., & Nathan, J. (1995). *A behavioral index of ethnocultural identity.* Unpublished manuscript, University of Hawaii, Honolulu.

Marsella, A. J., Horvath, A., & Tshushima, S. (1995, January). The multimodel measurement of ethnocultural identity. Paper presented at the meeting of the Hawaii Psychological Association, Honolulu, HI.

Marsella, A. J., & Kameoka, V. A. (1988). Ethnocultural issues in the assessment of psychopathology. In S. Wetzler (Ed.), *Measuring mental illness: Psychometric assessment for clinicians* (pp. 229–256). Washington, DC: American Psychiatric Press.

Marsella, A. J., Kinzie, D., & Gordon, D. (1973). Ethnocultural variations in the expression of depression. *Journal of Cross-Cultural Psychology, 4,* 435–458.

Marsella, A. J., Sartorius, N., & Jablensky, A. (1985). Depression across cultures. In A. Kleinman & B. Good (Eds.), *Culture and depression* (pp. 299–324). Berkeley: University of California Press.

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: A Journal of Consulting and Clinical Psychology, 2,* 114–121.

Mollica, R. F., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., & Lavelle, J. (1992). The Harvard Team Questionnaire: Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *Journal of Nervous and Mental Disease, 180,* 111–116.

Mollica, R. F., Wyshak, G., de Marneffe, D., Khuon, F., & Lavelle, J. (1987). Indochinese versions of the Hopkins Symptom Checklist-25: A screening instrument for the psychiatric care of refugees. *American Journal of Psychiatry, 144,* 497–500.

Newman, E., Kaloupek, D. G., & Keane, T. M. (in press). Assessment of PTSD in clinical and research settings. In A. C. McFarlane, B. van der Kolk, & L.

Weisaeth (Eds.), *Comprehensive text on posttraumatic stress*. Cambridge: Cambridge University Press.

Penk, W. E., & Allen, I. M. (1991). Clinical assessment of post-traumatic stress disorder (PTSD) among American minorities who served in Vietnam. *Journal of Traumatic Stress*, *4*, 41–66.

Prins, A., Kaloupek, D. G., & Keane, T. M. (1995). Psychophysiological evidence for autonomic arousal and startle in traumatized adult populations. In M. J. Friedman, D. Charney, & A. Deutch (Eds.), *Neurobiological and clinical consequences of stress: From normal adaptation to PTSD* (pp. 291–314). New York: Raven Press.

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

Sutker, P. B., Uddo-Crane, M., & Allain, A. N. (1991). Clinical and research assessment of posttraumatic stress disorder: A conceptual overview. *Journal of Consulting and Clinical Psychology*, *3*, 520–530.

Weathers, F. W., Blake, D. D., Krinsley, K. E., Haddad, W., Huska, J. A., & Keane, T. M. (1992, November). *The Clinician-Administered PTSD Scale: Reliability and construct validity*. Paper presented at the 26th Annual Convention of the Association for Advancement of Behavior Therapy, Boston.

Weathers, F. W., Litz, B. T., Keane, T. M., Herman, D. S., Steinberg, H. R., Huska, J. A., & Kraemer, H. C. (1993). *The PTSD Checklist: Description, use, and psychometric properties*. Unpublished manuscript.

Wilson, J. P., & Raphael, B. (Eds.). (1993). *International handbook of traumatic stress syndromes*. New York: Plenum Press.