DESNOS in Three Postconflict Settings: Assessing Cross-Cultural Construct Equivalence

Joop T.V.M. de Jong,[1,2,4] Ivan H. Komproe,[1] Joseph Spinazzola,[3] Bessel A. van der Kolk,[3] and Mark H. Van Ommeren[1]

This study examined the cross-cultural construct equivalence of the Structured Interview for Disorders of Extreme Stress (SIDES), an instrument designed to assess symptoms of Disorders of Extreme Stress Not Otherwise Specified (DESNOS). Participants completed the SIDES as a part of an epidemiological survey conducted between 1997 and 1999 among survivors of war or mass violence in Algeria (*n* = 652), Ethiopia (*n* = 1,200), and Gaza (*n* = 585). Findings indicated that the factor structure of the SIDES across samples was not stable; thus construct equivalence was not shown. A multistep interdisciplinary method is proposed to improve the cross-cultural construct validity of a psychiatric concept. This method accommodates universal chronic sequelae of extreme stress and accommodates culture-specific symptoms across a variety of cultures.

Complex posttraumatic stress disorder (PTSD) or Disorders of Extreme Stress Not Otherwise Specified (DESNOS; Herman, 1992) has been proposed as an alternative to Axis II personality disorder diagnosis when extreme trauma compromises the fundamental sense of self and relational trust at critical developmental periods (Roth, Newman, Pelcovitz, van der Kolk, & Mandel, 1997). The World Health Organization has also recognized the occurrence of such posttraumatic alterations and included in its ICD-10 classification a diagnostic category of "lasting personality changes following catastrophic stress," which comprises "impairment in interpersonal, social and occupational functioning," including "a hostile and mistrustful attitude towards the world, social withdrawal, feelings of emptiness and hopelessness, a chronic feeling of being 'on the edge' and constantly threatened and chronic sense of estrangement" (WHO, 1992, pp. 232–233).

To date, the research on DESNOS has primarily been directed towards measurement and validation of the DESNOS construct (e.g., Zlotnick & Pearlstein, 1997), examination of its clinical correlates and effects on associated psychopathological conditions and treatment outcome (e.g., Ford & Kidd, 1998), and establishment of preliminary guidelines for effective intervention for patients who have DESNOS (e.g., Luxenberg, Spinazzola, Hidalgo, Hunt, & van der Kolk, 2001). The accumulation of this research since the 1990s has provided strong support for the construct validity of DESNOS in Western samples of people who have a history of childhood intrafamilial abuse, but the precise prevalence rates of this disorder in diverse trauma populations is still unknown. Thus far, only one study that examined the DESNOS construct in a non-Western sample (Weine et al., 1998) has been published. In the study, current prevalence of DESNOS among 24 adolescent and adult Bosnian refugees exposed

[1]Transcultural Psychosocial Organisation (TPO), Amsterdam, The Netherlands.
[2]Vrije Universiteit, Amsterdam; Boston University School of Medicine, Boston, Massachusetts.
[3]The Trauma Center, Boston University School of Medicine; National Child Traumatic Stress Network, Boston, Massachusetts.
[4]To whom correspondence should be addressed at Transcultural Psychosocial Organisation, Keizersgracht 329, 1016 EE, Amsterdam, The Netherlands; e-mail: jdejong.tpo@pom.nl.

to "ethnic cleansing" was assessed approximately 1 year after their resettlement in the United States. None of the individuals assessed was found to suffer from all six symptom clusters of DESNOS. However, several methodological limitations such as small sample size restrict its generalizability to the prevalence of DESNOS in cross-cultural trauma settings.

Other than this study, research on DESNOS has been limited to selected and mainly treatment-seeking samples living in the West. The DESNOS construct was meant to address specifically the psychiatric sequelae of interpersonal victimization, particularly those forms of victimization that occur early in the life cycle and involve multiple traumatic events and extended exposure (Luxenberg, Spinazzola, & van der Kolk, 2001). These types of traumatic events are rampant in situations in which people are exposed to war, genocide, persecution, political repression, torture, ethnic cleansing, terrorism, poverty, and other calamities (de Jong, 2002; Green et al., 2003). The validity of the DESNOS construct, however, has never been evaluated in communities exposed to multiple prolonged extreme stressors.

The validity of a diagnostic concept cannot be taken for granted. According to Kendell and Jablensky (2003), it is important to determine the validity and utility of a diagnostic instrument for diagnoses because diagnostic categories defined by their syndromes should be regarded as valid only if they have been shown to be discrete entities that have natural boundaries that separate them from other disorders. Diagnostic categories that are not valid may, however, be useful as working concepts for clinicians because they provide information about outcome, treatment response, and cause. The current study evaluates the cross-cultural measurement equivalence of the Structured Interview for Disorders of Extreme Stress (SIDES) and the DESNOS concept (Pelcovitz et al., 1997). Without fulfillment of the requirement of measurement equivalence, interpretations of differences in scores may lead to erroneous conclusions in cross-cultural settings (Poortinga, 1975). Measurement equivalence is based on the concept of construct or theoretical validity, which is defined as the correlation of an observed variable with some theoretical construct (latent variable) of interest (Irvine & Carroll, 1980). It refers to the equivalence of theoretical validities across populations. Evidence for measurement equivalence is a psychometric prerequisite for the comparison of prevalence rates or mean scores of (sub)scales.

After assessing the measurement equivalence we discuss and interpret the results of our analyses with additional methodological requirements. Applying the concepts of content, criterion, technical, conceptual, and semantic equivalence illustrates how culture affects the scores on the SIDES in our data set (Flaherty et al., 1988; de Jong & Van Ommeren, 2002).

Finally we propose a multistep procedure to improve the cross-cultural construct validity of the DESNOS concept. This long-term process implies an interdisciplinary methodological approach to the development of a psychiatric concept and the construction of an instrument that can be used across cultures.

Method

Samples

Details of both the sampling and the populations' local contexts have been published (de Jong et al., 2001; de Jong, Komproe, & Van Ommeren, 2003). In short, the populations are in three different regions of the world: North Africa (Algeria, settled community affected by recent terrorism [Aït Sidhoum, Arar, Bouatta, Khaled, & El Masri, 2002]), East Africa (Ethiopia,refugee camps that have a very low-quality living environment [Aptekar & Giel, 2002]), and the Middle East (Gaza, long-term refugee population affected by ongoing conflicts studied between the first and the second Intifada [Qouta & El Sarraj, 2002]). The countries were selected on the basis of three criteria: (1) the presence of intervention programs started or supported by the Transcultural Psychosocial Organization, a World Health Organization (WHO) Collaborating Centre for Refugees and Ethnic Minorities; (2) the absence of ongoing high-intensity conflict at the time of data collection; and (3) the availability of local human resources to organize and conduct data collection.

As described in detail previously (de Jong et al., 2001, 2003), multistep random sampling procedures were used to select representative samples in the countries. By using random sampling, addresses were selected from lists of resident addresses that had been obtained from local authorities. One person was randomly selected from each selected address. The Algerian sample ($n = 653$) was randomly selected from the Gouvernorat d'Algiers. The Ethiopian sample ($n = 1,200$) was randomly selected from temporary shelters for refugees from Eritrea. The Palestinian sample ($n = 585$) was randomly selected from refugee camps, cities, and resettlement areas in Gaza.

Characteristics of the samples have been described previously (de Jong et al., 2001, 2003).

Measures

DESNOS symptoms were assessed by the SIDES interview (Pelcovitz et al., 1997). The DESNOS interview

under study consisted of six clusters: (1) alteration in regulation of affect and impulses (six items), (2) alterations in attention or consciousness (two items), (3) alterations in self-perception (six items), (4) alterations in relations with others (three items), (5) somatization (five items), and (6) alterations in systems of meaning (two items). The seventh cluster, alterations in perception of the perpetrator, was left out of the analyses in accordance with the suggestions of Pelcovitz and others (1997).

In all countries the process of translation and adaptation for each setting involved (1) examination of the instruments by bilingual experts, (2) translation into local language, (3) literal back translation by different translators, (4) examination of the back translation by monolingual experts, (5) back translation of all items changed by the monolingual experts, (6) examination of the back translation by a bilingual group informed by the discussion of the monolingual experts, and (7) pilot testing.

Procedures

Local field-workers who had received 2 weeks of training in the administration of the instruments by an experienced trainer administered the instrument. Verbal rather than written informed consent was obtained because of illiteracy and fear of signing forms. Local program directors, their boards, and local authorities approved the research procedures, which were consistent with the Declaration of Helsinki (Forty-Eighth World Medical Assembly, Declaration of Helsinki, 1997).

Analyses

Measurement equivalence was evaluated statistically by testing the factorial invariance of the items of the SIDES interview across different samples. We used both exploratory factor analyses and multisample confirmatory factor analyses for testing the factorial invariance of the SIDES subscales.

First, we performed exploratory factor analyses of the six subsales of SIDES in the data sets. The scores on the six subscales were centered for each respondent, to obtain similar ranges of variances of scores on the subscales and reduce multicollinearity between the variables. These z-scores of the subscales of the SIDES were entered as variables in the factor analyses. We used principal axis factoring as the extraction method, because we were interested in the explanation of the covariances between the subscales, rather than of the variances of the subscales. These factor analyses yielded the most likely factor struc-

ture of the six subscales of SIDES in the three data sets. A variable was assigned to a factor and used for factor labeling if its loading on that respective factor was greater than .35 and its loading on any other factor less than .35 (Nunnally, 1978).

Second, we compared the obtained factor structures of the subscales of SIDES of the different samples. We divided the Ethiopian sample into two equal subsamples by means of a random allocation of respondents into one of the two groups ($n1 = 600$; $n2 = 600$). The $n1$ sample was used as the reference sample in the cross-validation procedure. This reference sample was compared with the $n2$ Ethiopian sample and the samples from Algeria and Gaza. The division of a large sample in a cross-validation procedure is superior to the conventional single-group confirmatory factor-analytic strategy because it minimizes capitalization on chance in two ways: (1) by using multiple samples, thereby allowing cross-validation of obtained models (Cudeck & Browne, 1983), and (2) by using a specified model (e.g., found in sample $n1$) instead of data-driven models (McCallum, Roznowski, & Necowitz, 1992).

We used LISREL 8 for multisample confirmatory factor analyses to evaluate equality of factor structures (Jöreskog & Sörbom, 1993). Confirmatory factor analysis involves testing a series of hypotheses about the interview's measurement properties across groups. If these analyses indicate that the measurement is invariant across cultural groups, quantitative comparisons of scale scores may be justified. An invariant pattern of factor loadings of the SIDES subscales across different samples is an indication of the stability of the factor structure. Replication of a factorial composition derived from a reference sample (e.g., the $n1$ Ethiopian sample) in another sample is an indicator of construct validity (Devins et al., 1988). Similarity of patterns of factor loadings can be defined on different levels; thus there are different hypotheses to test the similarity of factorial composition (Bollen, 1991; Jöreskog & Sörbom, 1993). We tested the different hypotheses of factorial invariance by comparing the absolute fit of different factor models (Bryne, Shavelson, & Muthén, 1989). In this study we distinguished the following hierarchical models: (1) a model in which the number and pattern of factors are equal across samples (model A); (2) model A with the additional constraint that the factor loadings are equal across samples (model B); (3) model B with the additional constraint that the error variances are equal across samples (model C); (4) model C with the additional constraint that the covariance matrices of factors are equal across samples (model D). The difference in fit between (1) model A and model, B, (2) model A and model C, (3) model A and model D was computed.

If the difference in chi square between models was not significant, equality of factor structure was assumed.

The χ^2 test is a global test of a model's ability to reproduce the sample variance/covariance matrix (Jöreskog & Sörbom, 1993). A nonsignificant χ^2 indicates high congruence between model and data, but because of the power of the test even small residual variances produce a significant χ^2 in large samples (Fornell, 1983). The degree of dissimilarity between factor structures across samples determines the difference in the χ^2 between both test models. For the comparison of different models we are not specifically interested in the absolute fit of the model, but in the increment in fit of the model compared to that of others. When the difference in the χ^2 value of models, $\Delta\chi^2$, is not significant, the hypothesis of invariant factor loadings is tenable (Jöreskog & Sörbom, 1993). This procedure is preferable to the evaluation of the absolute χ^2-goodness-of-fit index because the latter is substantially affected by sample size (Marsh, Balla, & McDonald, 1988).

Differences between percentages of symptoms of DESNOS of samples were calculated by using the algorithm $|t| = Sqrt[(N1 \times N2)/(N1 + N2)] \times |p1 - p2|/Sqrt(p \times q)$, where $p = (p1 \times N1 + p2 \times N2)/(N1 + N2)$ and $q = 1 - p$.

Results

Analyses of the sociodemographic characteristics of the three samples revealed some dissimilarities among the samples. The sex ratio varied across samples, χ^2 (3) = 37.4, $p < .05$. Compared to that in Algeria, there are higher proportions of women in the samples of Ethiopia and Gaza. The mean age varied across samples: Algeria (41 years), Ethiopia (34 years), and Gaza (32 years), F (3, 3197) = 179.8, $p < .001$. All age comparisons between pairs of samples were significant according to the Scheffé post hoc test ($p < .05$). In Algeria 48% of the respondents were married and 46% not married; in Ethiopia 36% of the respondents were married, 36% widowed or separated, and 21% not married; in Gaza 67% of the respondents were married and 31% not married, χ^2 (8) = 556.5, $p < .001$. On average respondents were more highly educated in Gaza (62% at least secondary school) than in Algeria (39%) and Ethiopia (31%), χ^2 (8) = 671.1, $p < .001$. Virtually all respondents in Algeria and Gaza (96%) were Muslim, whereas in Ethiopia 93% were Orthodox, χ^2 (2) = 93.6, $p < .001$.

Table 1 shows that the factor loadings of the six scales on the latent factor vary across samples. When factor loadings were squared, the obtained communalities refer to the explained variance of the subscales by the latent factor. The explained variance of the subscales of DESNOS varied from 25% to 69% in the Ethiopian $n1$ sample, from 36% to 57% in the Algerian sample, from 32% to 61% in the Ethiopian $n2$ sample, and from 54% to 72% in the sample from Gaza.

The variation of factor loadings refers also to different interrelationships (covariances) among the subscales. This finding implies that although one overall factor was found responsible for the scores on the subscales, the constitution of this factor may differ across samples.

Table 2 summarizes the results of the tests of equality of factor structures. As the p values of the χ^2 differences in fit of the factor models in the two samples of Ethiopia are all nonsignificant, one can conclude that the factor structure is invariant across the two samples. Therefore, the interrelationships of the six subscales are the same in both samples; thus the constitution of the construct DESNOS is invariant in the samples from Ethiopia.

All p values of the χ^2 comparisons of the factor structure models between the Ethiopian reference sample (Ethiopian $n1$ sample) and the samples from Algeria and Gaza were significant at the .05 level. The chi-square differences ($\Delta\chi^2$) between the reference sample and the Algerian sample ranged from 3.26 (model A against model D) to 185.14 (model A against model C). The χ^2 comparison of the reference sample and the Gaza sample

Table 1. Factor Loadings of the Six Subscales of the Disorders of Extreme Stress Not Otherwise Noted Interview in Samples From Ethiopia, Algeria, and Gaza

Scale	Ethiopia $n1$ $n = 600$	Ethiopia $n2$ $n = 600$	Algeria $n = 652$	Gaza $n = 585$
Alterations in regulation of affect and impulses	.70	.73	.76	.86
Alterations in attention or consciousness	.68	.67	.67	.81
Alterations in self-perception	.82	.78	.69	.81
Alterations in relations with others	.62	.58	.67	.83
Somatization	.56	.62	.61	.78
Alterations in system of meaning	.63	.66	.70	.75
Explained variance of factor	53.8%	54.5%	55.6%	70.9%

Table 2. Test of the Equality of Factor Structures of Disorders of Extreme Stress Not Otherwise Noted Interview Among Samples From Ethiopia, Algeria, and Gaza

Model	$\chi^2_{Ethiopia1}$	$\chi^2_{Ethiopia2}$	$\Delta\chi^2$	Δdf
Model A	346.83	371.68	—	
Model B	350.82	375.57	7.86	5
Model C	353.37	377.47	4.46	6
Model D	353.41	377.59	0.16	1

Model	$\chi^2_{Ethiopia1}$	$\chi^2_{Algeria}$	$\Delta\chi^2$	Δdf
Model A	346.83	411.16	—	
Model B	363.12	427.06	32.18	5*
Model C	473.96	483.36	185.14	6*
Model D	485.71	492.87	3.26	1

Model	$\chi^2_{Ethiopia1}$	χ^2_{Gaza}	$\Delta\chi^2$	Δdf
Model A	346.83	651.76	—	
Model B	382.57	665.28	49.26	5*
Model C	404.58	709.27	115.26	6*
Model D	412.42	719.68	18.25	1*

Note. Models of invariance: Model A = Model in which the number and pattern of factors are equal across samples; Model B = Model A with the additional constraint that the factor loadings are equal across samples; Model C = model B with the additional constraint that the error variances are equal across samples; Model D = model C with the additional constraint that the covariance matrices of factors are equal across samples.
$\chi^2_{Ethiopia1, Ethiopia2, Algeria, Gaza}$ = χ^2 value of tested factor structure (model) in sample; χ^2tot = total χ^2 value of model with the same factor structure in two different samples; $\Delta\chi^2$ = χ^2 difference between two hierarchical models of invariance; Δdf = difference in degrees of freedom between two models of invariance.
*$p < .01$.

showed a $\Delta\chi^2$ ranging from 18.25 (model A against model D) to 115.26 (model A against model C).

These findings indicate that the factor structure across samples from different countries is not stable, so that the constitution of the theoretical concept (latent factor) is not the same in the different samples.

Table 3 shows the percentages of present symptoms of DESNOS in the three data sets and the percentages of the symptoms in the *Diagnostic and Statistical Manual of Mental Disorders,* fourth edition (DSM-IV), PTSD field trial. We compared the percentages of our data set with the total sample percentages (proportional average percentage of Early Onset and Late Onset samples) of the DSM-IV field trial data (Pelcovitz et al., 1997).

Discussion

Interpreting the Results With Additional Culturally Appropriate Methodological Requirements

In our study, a number of symptoms were seldom endorsed in the participating countries, when compared

to those endorsed in the DSM field trial, even though we used a careful multistep procedure to translate our instruments (Van Ommeren et al., 1999). We comment on these results with the help of the different forms of equivalencies described by Flaherty and associates (1988) and with the notion of social desirability, which we regard as one aspect of technical equivalence. Our comments are limited to those items that are significantly less prevalent in our study than in the DSM-IV field trial and that in our view are interesting from a cultural point of view. On scale I, the prevalence of suicidal ideation in our study was almost nine times lower than in the DSM field trial. This result may reflect technical inequivalence (stemming from social desirability) because suicide is taboo in both the Islamic and Coptic religions; in Buddhism it is considered to result in numerous reincarnations as an inferior being, such as an animal. Scores on risk taking were 10 times lower than in the field trial and may result from a lack of content equivalence (i.e., the content of the item is relevant to the phenomena of each culture being studied; Flaherty et al., 1988). When a respondent is permanently surrounded by dangers such as the random terrorist attacks in Algeria, and when these external events fit with a psychological tendency to attribute events to external causes, the events may shadow or overrule a tendency to take risks and result in an underreporting of risk taking. In other words, the second part of the related question on the SIDES ("Did you expose yourself to places that are not safe?") has little relevance in a context where lack of safety is a daily reality and where norms about safety are accommodated to the harshness of the context. On scale II, amnesia scores were four times lower than in the field trial because of a lack of technical equivalence (i.e., the comparability of the performance of the assessment method). The relevant SIDES item consists of three subquestions that need reformulation and are difficult to answer, especially because they are lumped together in one long phrase ("Do you find that there are parts of your life you cannot remember, or are you confused about what happened, or are you unsure whether certain important things did or did not happen to you?"). In non-Western cultures these three subquestions readily provoke concretisms such as "As for most people, there are parts of my life I cannot remember, so you are never sure whether things did or did not happen to you"—a statement applicable to all human beings.

In scale III, the questions about guilt and responsibility and shame both scored more than six times lower than in the DSM field trial. In our view both a lack of conceptual equivalence (i.e., the instrument measures the same theoretical construct in each culture) and a lack of semantic equivalence (i.e., the meaning of each item is the

Table 3. Percentage Endorsement of Structured Interview for Disorders of Extreme Stress Items in Algeria, Ethiopia, and Gaza and in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Posttraumatic Stress Disorder Field Trial

Scale	Algeria $n = 652$ %	Ethiopia $n = 1200$ %	Gaza $n = 585$ %	Field Trial $n = 236$ %
1. Alterations in regulation				
of affect and impulses	35.2	13.4	17.6	NA
Affect	44.3	14.1	21.9	72.0
Anger	32.5	17.4	16.2	68.0
Self-destructive	32.0	24.2	17.1	49.5
Suicidal[a]	6.9	5.9	3.4	53.0
Sexual involvement	30.0	60.1	23.6	74.0
Risk taking[a]	5.7	1.2	3.8	40.0
2. Alterations in attention				
or consciousness	56.2	23.0	28.4	NA
Amnesia[a]	19.0	10.8	9.7	62.0
Dissociative[a]	52.2	18.3	26.5	69.5
3. Alterations in				
self-perception	31.1	16.9	15.0	NA
Ineffectiveness	30.3	13.0	13.7	47.0
Damage	15.2	9.2	11.8	62.0
Guilt and responsibility[a]	12.6	7.3	4.4	59.0
Shame[a]	8.3	8.3	5.6	49.5
Understand	32.5	14.5	12.1	68.5
Minimize	6.9	10.4	3.9	25.5
4. Alterations in relations				
with others	66.3	30.3	25.0	NA
Trust	64.0	28.8	23.1	84.5
Revictimization	22.1	2.7	8.2	46.0
Victimizing others[a]	3.8	1.2	0.9	18.0
5. Somatization	54.4	16.5	22.9	NA
Digestive	61.3[b]	6.3	29.2	65.0
Chronic pain	44.0[b]	31.3	23.4	49.0
Cardiopulmonary	35.5	10.3	11.6	65.5
Conversion	33.7	11.2	11.6	42.0
Sexual[a]	17.0	5.6	6.8	72.5
6. Alterations in system				
of meaning	65.1	46.9	28.2	NA
Hopelessness	58.5	34.3	23.9	69.5
Beliefs	53.6[b]	39.9	22.4	59.5
Disorders of extreme				
Stress	13.2	2.2	5.6	50.6

[a]Differences in scores that are discussed in the text.
[b]Percentage not significant different from ($p > .05$) field trial percentage.
[c]Percentage of subscale not available.

same in each culture after translation) may be the cause. The related SIDES questions ("Do you feel chronically guilty about all sorts of things?" and "Do you feel too ashamed of yourself to let people get to know you? Did you hide from others? Did you make up a cover story?") probe for the abstract psychological concepts guilt and shame, which require distancing from the self. However, in the cultures involved, guilt and shame are related to actual behavior that is conceived as right or wrong in the case of guilt and of actual social or living conditions that generate shame such as being poor or walking around in ragged clothes. In addition, the three subquestions to probe for shame may be confusing and hence may also

lack technical equivalence. Similarly, in scale IV, the question about victimizing others, which scored 10 times lower than in the field trial, evokes a socially desirable answer because admitting that you hurt somebody as you were hurt, would create shame and hence is difficult to admit. In scale V the question about sexual somatizations (for males, burning sensation in sexual organs and impotence; for females, burning sensation in sexual organs, irregular menstrual periods, excessive premenstrual tension or menstrual bleeding) scored nine times lower than in the field trial. This finding is probably also related to cultural inhibition and thus to social desirability and hence a matter of technical equivalence.

In our opinion the questions about dissociation raise some interesting questions and show a lack of content equivalence. For example, a question about "keeping track of time" has little relevance in three of the four cultures being studied because they have their own calendar and many respondents do not have a watch. The question about "spacing out when you feel frightened or under stress" creates a problem with semantic equivalence because this dissociative experience is very hard to translate (even in most Western European languages) and anything getting close to "spacing out" can easily be confused with spirit possession of the person in non-Western cultures, as opposed to the association between spacing out and drugs in the West. Similarly, another aspect of dissociation, assessed by asking whether "you sometimes feel like there are two people living inside you who control how you behave at different times," creates a problem of conceptual equivalence. It requires a focus on cognitive introspection; outside the West cognitive focus on relationships with various others is more likely. Even healers and shamans who enter a trance many times a day would not be likely to agree that it "feels like there are two people living inside them who control their behavior," because during possession the spirit controls one's behavior. This problem is even more likely for an average respondent who attributes a dissociative experience to an ancestor, a deity, or a ghost. The answers to the questions about dissociation gave a mixed picture. In Algeria the score was only somewhat lower than in the field trial, and in Ethiopia and Gaza scores were considerably lower. In Algeria spirit possession is a common phenomenon both as an idiom of distress and within a complementary/alternative healing setting and therefore may be differentiated more easily, even though it may be attributed to an external force.

Some Explanations for the Limited Cross-Cultural Applicability of the SIDES

For a comparison of prevalence of DESNOS across countries, the SIDES appears to be an instrument that requires adaptation to other cultural settings. We provide several explanations for this inadequacy by looking at aspects of reliability and validity. McHugh and Vallis (1986) defined *reliability* as "verification of observations"; it is "the consistency with which one can make an observation . . . which is demonstrated by the correlation between the results of observers using the same technique to make that observation." Although we tried to increase reliability by arranging that one trainer in the participating sites taught the use of the questionnaire, we did not establish interrater reliability.

Apart from reliability, we question the cross-cultural validity of the SIDES. Validity is the "verification of assumptions," that is, the verification of the psychiatric categories themselves (Kleinman, 1988). Psychiatrists can be trained to make the same observations, but we cannot be sure that the observations are valid, that is, whether a patient does or does not have an abnormal mental state. For example, a Palestinian militant may qualify for suicidal ideation on the first DESNOS subscale (Alteration in regulation of affect and impulses). Interviewers may determine with high reliability how some Palestinians score positively on this DESNOS subscale. But the determination of whether these reports refer to an abnormal mental state is an interpretation. This interpretation is related to knowledge of the group's behavioral norms and what they consider as normal or deviant experiences or reactions.

To assess whether the current SIDES has cross-cultural validity by accommodating the range of deviant reactions known as DESNOS in different cultures, we take a closer look at the development of the SIDES. The universal aspiration of the development of the DESNOS concept undermines its cross-cultural applicability. First, symptoms were generated by a review of predominantly Western literature. Second, after consulting more than 50 authorities on traumatic stress, a panel of 12 researchers—most of them Caucasians—agreed on a final list of 27 items (Pelcovitz et al., 1997). Subsequently, the DSM field trial interviewed 523 Americans, 80% of whom were white (van der Kolk et al., 1986). Despite this careful and thorough methodology, three factors contributed to a lack of cross-cultural validity: First, at the time of the literature review, there existed a lack of qualitative and quantitative data from non-Western cultures regarding the sequelae of childhood sexual abuse, rape, or spouse abuse. Second, the panel of researchers who decided which items were included in the SIDES captured the experience of their predominantly Western patients in a Western setting. Validity is an invariate characteristic of a diagnostic category and is not dependent on context (Kendell & Jablensky, 2003). However, without verifying its variability or stability in different cultural settings one can never conclude that the construct is invariate and thus not dependent on context. Third, the procedure sought variety to the detriment of what Kendell and Jablensky (2003) call "zones of rarity" or discrete diagnostic entities that have natural boundaries.

Multistep Procedure to Improve Cross-Cultural Construct Validity

A range of DESNOS symptoms can be found or elicited in different cultures. However, this study does not

confirm the existence of a predefined universal diagnostic construct or syndrome. A careful phenomenological approach during the development stage of the DESNOS construct might have yielded unknown (sub)types or variations of DESNOS that were not captured by the procedures described.

We propose a procedure for the construction of a universal core module to capture the consequences of extreme stress across cultures, with local modules that fit culture-specific expressions of extreme stress. To develop a cross-culturally useful diagnostic construct we propose to distinguish type A, type B, and type C symptoms. Type A or core symptoms are those that are the same in all cultures. These symptoms (e.g., difficulties modulating anger as a universal symptom of the universal process affect dysregulation) form the universal items of the instrument (the core module). These symptoms are likely to be based on universal neurobiological processes and on molecular genetic and molecular biological, neurochemical, neurophysiological, and cognitive neurological characteristics, and may be considered defining characteristics (Andreasen, 2001; Kendell & Jablensky, 2003). Type B symptoms are those that are unique to a culture but that reflect universal underlying problems (e.g., self-injury as a cultural symptom of the universal process affect dysregulation). Type B symptoms should be incorporated in the local module. Type C symptoms are expressions of culture-specific processes that have specific symptoms (e.g., low self-esteem as a culture-specific symptom of the culture-specific process of disturbances in self-perception). Type C symptoms should ideally also be incorporated in the local module.

Most authors agree that universal characteristics in the presentation of psychopathological conditions are more prominent when biophysiological factors play a prominent role, and that various other idioms of distress and illness factors are culture-specific and may be better explained within a sociocultural paradigm (de Jong & Van Ommeren, 2002). Therefore, the first step must elicit invariable dimensions involved in neurobiological processes that evolved in the course of the evolutionary history of primates. It is likely that a limited number of type A or core symptoms will get us closer to a universal construct. There are several possibilities, for example, to study the neurobiological substrate of (1) adult survivors of type I trauma in childhood, (2) adult survivors of type II trauma in childhood, (3) adult survivors of type I trauma in adulthood, and (4) adult survivor of type II trauma in adulthood. The second step could be to deconstruct syndromes that have been used in psychiatry over time (e.g., battered women syndrome, concentration camp syndrome, torture syndrome, shell shock, PTSD,

DESNOS). The third step would be to compare all data sets on WHO's category "lasting personality changes following catastrophic stress" with the data on DESNOS. The second and third steps may suggest type A, B, and C symptoms because torture has been practiced and described in many cultures, whereas the literature on shell shock or battered women is limited to historical episodes and geographic areas. Subsequently, one could deconstruct those DESNOS dimensions that belong to (other) Axis I or II disorders and that contribute to fuzzy boundaries between DESNOS and other disorders. For example, somatization is characteristic of somatoform disorder, depressive disorder, anxiety disorder, and culture-specific illness behavior and contaminates the DESNOS syndrome. Likewise, anger and self-destructiveness also are characteristic of borderline personality disorder, and anger and victimizing of others of antisocial personality disorder. Similarly, the second step may include a review and an analysis of qualitative and quantitative data from non-Western cultures regarding the sequelae of childhood sexual abuse, rape, or spouse abuse that have been collected since the origin of the DESNOS construct. There are several techniques to distinguish between discrete entities and continuous variation such as, a combination of discriminant function analysis and admixture analysis or a "grade of membership" model (Kendell & Jablensky, 2003).

The third step aims at finding type C symptoms in a variety of cultures and settings by means of qualitative techniques. Focus groups can be used as a way of socially generating expressions of culture-specific consequences of extreme stress. Focus groups should be organized among both urban and rural people (e.g., office workers and peasants); among patients, healthy persons, mental health professionals, and traditional healers; and among high-risk groups (e.g., former combatants and survivors of human rights violations). Because the data yielded by focus groups can be the product of social dynamics within the group, the data from several focus groups sampled from the same population should be compared to protect against this possibility. Upon completion of the focus group work, a panel of the local investigator and the focus group moderator should analyze the focus group transcripts to describe the type C symptoms in that culture and add them to the developing instrument (cf. WHOQOL group 1995, de Jong & Van Ommeren, 2002).

In our view this combination of empirical research techniques with culturally sensitive research strategies would be a step forward to develop a universal psychiatric concept of disorders of extreme stress while accommodating culture-specific symptoms across a variety of cultures and continents.

References

Aït Sidhoum, M.A., Arar, F., Bouatta, C., Khaled, N., & El Masri, M. (2002). Terrorism, traumatic events and mental health in Algeria. In J.T.V.M., de Jong (Ed.), Trauma, war and violence: Public mental health in socio-cultural context (pp. 376–405). New York: Kluwer/Plenum.

Andreasen, N.C. (2001). The validation of psychiatric diagnosis: New models and approaches [editorial]. American Journal of Psychiatry, 152, 161–162.

Aptekar, L., & Giel, R. (2002). Walks in Kaliti, life in a destitute shelter for the displaced in Addis Ababa. In J.T.V.M., de Jong (Ed.), Trauma, war and violence: Public mental health in socio-cultural context (pp. 337–367). New York: Kluwer/Plenum.

Bollen, K.A. (1991). Structural equations with latent variables. New York: Wiley Interscience Publication.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105, 456–466.

Cudeck, R., & Browne, M.W. (1983). Cross-validation of covariance structures. Multivariate Behavioral Research, 18, 147–167.

de Jong, J.T.V.M. (Ed.). (2002). Trauma, war and violence: Public mental health in socio-cultural context. New York: Kluwer/Plenum.

de Jong, J.T.V.M., Komproe, I.H., & Van Ommeren, M. (2003). Common mental disorders in post-conflict settings. Lancet, 361, 2128–2130.

de Jong, J.T.V.M., Komproe, I.H., Van Ommeren, M., El Masri, M., Araya, M., Khaled, N., et al. (2001). Lifetime events and posttraumatic stress disorder in four postconflict settings. Journal of the American Medical Association, 286, 555–562.

de Jong, J.T.V.M., & Van Ommeren, M.H. (2002). Toward a culturally informed epidemiology: A pragmatic model for qualitative and quantitative psychiatric research in cross-cultural contexts. Transcultural Psychiatry, 39, 422–433.

Devins, G.M., Orme, C.M., Costello, C.G., Binik, Y.M., Frizzel, B., Stam, H.J., et al. (1988). Measuring depressive symptoms in illness populations: Psychometric properties of the Center for Epidemiologic Studies Depression (CES-D) scale. Psychology and Health, 2, 139–156.

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

Ford, J.D., & Kidd, P. (1998). Early childhood trauma and disorders of extreme stress as predictors of treatment outcome with chronic posttraumatic stress disorder. Journal of Traumatic Stress, 11, 743–761.

Fornell, C. (1983). Issues in the application of covariance structure analysis. Journal of Consumer Research, 9, 443–448.

Forty-Eighth World Medical Assembly. (1997). Declaration of Helsinki: Recommendations guiding physicians in biomedical research involving human subjects. Journal of the American Medical Association 277, 925–926.

Green, B., Friedman, M., de Jong, J.T.V.M., Solomon, S., Keane, T.M., Fairbank, J.A., et al. (Eds.). (2003). Trauma interventions in war and peace: Prevention, practice, and policy. New York: Plenum-Kluwer.

Herman, J. (1992). Complex PTSD. Journal of Traumatic Stress, 5, 377–391.

Irvine, S.H., & Carroll, W.K. (1980). Testing and assessment across cultures: Issues in methodology and theory. In H.C. Triandis & J.W. Berry (Eds.), The handbook of cross-cultural psychology (Vol. 2, pp. 127–180). Boston: Allyn & Bacon,

Joreskög, K.G., & Sörbom, D. (1993). LISREL 8: Structural equation modelling with the SIMPLIS command language. Chicago: Scientific Software International.

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. American Journal of Psychiatry, 160, 4–12.

Kleinman, A. (1988). Rethinking psychiatry. New York: The Free Press.

Luxenburg, T., Spinazzola, J. Hidalgo, J., Hunt, C., & van der Kolk, B. (2001). Complex Trauma and the Disorders of Extreme Stress (DESNOS) diagnosis, part two: Treatment. Directions in Psychiatry, 11, 395–415.

Luxenberg, T., Spinazzola, J., & van der Kolk, B.A. (2001). Complex Trauma and Disorders of Extreme Stress (DESNOS) diagnosis: Part 1. Assessment. Directions in Psychiatry, 21, 373–394.

MacCallum, R.C., Roznowski, M., & Necowitz, L.B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin, 111, 490–504.

Marsh, H.W., Balla, J.R., & McDonald R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391–410.

McHugh, S., & Vallis, T.M. (Eds.). (1986). Illness behavior. New York: Plenum.

Nunnally, J.C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Pelcovitz, D., van der Kolk, B., Roth, S., Mandel, F., Kaplan, S., & Resick, P. (1997). Development of a criteria set and a structured interview for disorders of extreme stress (DESNOS). Journal of Traumatic Stress, 10, 3–16.

Poortinga, Y.H. (1975). Limitations on intercultural comparison of psychological data. Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 30, 23–39.

Qouta, S., & El Sarraj, E. (2002). Community mental health as practiced by the Gaza Community Mental Health Programme. In J.T.V.M., de Jong (Ed.), Trauma, war and violence: Public mental health in socio-cultural context (pp. 317–337). New York: Kluwer/Plenum.

Roth, S., Newman, E., Pelcovitz, D., van der Kolk, B., & Mandel, F. (1997). Complex PTSD in victims exposed to sexual and physical abuse. Journal of Traumatic Stress, 10, 539–555.

Van der Kolk, B.A., Pelcovitz, D., Roth, S., Mandel, F., McFarlane, A.C., & Herman, J. (1996). Dissociation, somatization and affect dysregulation: The complexity of adaptation to trauma. American Journal of Psychiatry, 153, 83–93.

Van Ommeren, M., Sharma, B., Thapa, S., Makaju, R., Prasain, D., Bhattari, R., et al. (1999). Preparing instruments for transcultural research: Use of the translation monitoring form with Nepali-speaking Bhutanese refugees. Transcultural Psychiatry, 36, 285–301.

Weine, S.M., Becker, D.F., Vojvoda, D., Hodzic, E., Sawyer, M., Hyman, L., et al. (1998). Individual change after genocide in Bosnian survivors of "ethnic cleansing": Assessing personality dysfunction. Journal of Traumatic Stress, 11, 147–153.

WHOQOL Group. (1995). The World Health Organization Quality of Life Assessment (WHOQOL): Position paper of the World Health Organization. Social Science and Medicine, 41, 1403–1409.

World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization.

Zlotnick, C., & Pearlstein, T. (1997). Validation of the Structured Interview for Disorders of Extreme Stress. Comprehensive Psychiatry, 38(4), 243–247.