Chapter 10

The Impact of Event Scale: Revised

Daniel S. Weiss

INTRODUCTION

Posttraumatic stress disorder (PTSD) was introduced into the world psychiatric nomenclature in 1978 (World Health Organization, 1978) with the publication of the ICD-9, documenting the cross-cultural recognition of the typical symptomatic response to exposure to traumatic life events (e.g., Horowitz, 1976). The characteristic core of the disorder includes the distressing oscillation between intrusion and avoidance. Intrusion is characterized by nightmares, unbidden visual images of the trauma or its aftermath while awake, intrusive thoughts about aspects of the traumatic event, sequelae, or self-conceptions. Avoidance is typified by deliberate efforts to not think about the event, not talk about the event, and avoid of reminders of the event. Also characteristic are more active attempts to push memories and recollections of the event or its aftermath out of mind by increasing use of alcohol or drugs, overworking, or other strategies designed to divert attention or to so exhaust someone that he or she is temporarily untouched by the intrusive phenomenology. In addition to the frank avoidance, Horowitz also described emotional numbing as a not uncommon sequel to exposure to a traumatic life event (Horowitz, 1975; Horowitz & Kaltreider, 1977). There is empirical evidence supporting three of these four phenomena. For example in analyses of the most commonly used structured clinical interview for PTSD, the Clinician-Administered PTSD Scale (Weathers, Keane, & Davidson, 2001), evidence of the prominence of these clusters has been presented (King, Leskin, King, & Weathers, 1998).

Following from this conceptualization, Horowitz and colleagues (Horowitz, Wilner, & Alvarez, 1979) published a simple but powerful self-report measure for assessing the magnitude of symptomatic response in the past 7 days to a specific traumatic life event that was titled the Impact

of Event Scale (IES). (Sadly, many citations and publications using the IES have used the plural of "event," labeling it the *Impact of Events [sic] Scale*. This innocent error, occurring perhaps because when the name of the scale is spoken one cannot easily distinguish between singular and plural, may have influenced bibliographic information and searches to some extent, since searching for either may overlook the other.)

Published before the appearance of the formal diagnostic criteria (American Psychiatric Association, 1980), the original IES comprised two subscales: intrusion (the sum of seven items), and avoidance (the sum of eight items) that mapped on to what was described in the B and C criteria of the diagnosis of PTSD – the signs and symptoms of intrusive cognitions and affects together or oscillating with periods of avoidance, denial, or blocking of thoughts and images. The scale used a somewhat unusual response format: Not at all = 0, Rarely = 1, Sometimes = 3, and Often = 5.

The scale did not assess the third set of PTSD symptoms, the hyperarousal symptoms presented in the D criterion of the diagnosis of PTSD with the exception of disturbances in sleep. Thus, the phenomena of hypervigilance, angry outbursts, and exaggerated startle response were not covered in the original scale. The findings from the Department of Veterans Affairs Cooperative Study that examined in the laboratory the hyperarousal phenomena of PTSD (Keane et al., 1998) found, somewhat unexpectedly, that a proportion of those diagnosed with PTSD did not show the characteristic psychophysiological arousal to laboratory triggers. Despite the steady increase in the evidence that the fight or flight response manifest in the hypothalamic–pituitary–adrenal (HPA) axis is clearly implicated in the development, maintenance, or both of PTSD (Deebiec & Ledoux, 2006; Pitman et al., 2002; Yehuda, 2006), the results of the VA Cooperative Study are consistent with the notion that a subset of those with PTSD have less salient evoked hyperarousal responses. The exclusion of this domain from the original IES may have been associated with this phenomenon, but as a consequence the original IES was unable to assess symptomatic status in the three domains that comprise the diagnosis of PTSD.

In support of the two subscales, Horowitz et al. (1979) presented data that were consistent with two homogeneous clusters of items tapping by intrusion and avoidance (Cronbach's, 1951; alpha for intrusion = 0.79, for avoidance = 0.82). The correlations between the two subscales ($r = 0.42$, 18% of the variance) were small enough to allow for meaningful independence of the two subscales. The test–retest reliability was satisfactory, with coefficients of 0.87 for intrusion and 0.79 for avoidance.

Zilberg, Weiss, and Horowitz (1982) showed that the psychometric characteristics of the IES and the accompanying conceptual model of responses to traumatic stress that had given rise to its development held in a cross-validation sample. The data were from outpatients with

traumatic grief and a contrast group who had also experienced the death of a parent but had not sought treatment and were adjusting normally to their loss. Both groups were evaluated at three points in time.

The results revealed that item endorsement percentages ranged from 44% to 89% when the patient and contrast groups were combined, underscoring the salience of the phenomena tapped by the item pool. The rank order of items based on frequency of endorsement in the combined group was put side by side with the rank order reported in the initial publication of the IES. The Spearman rank correlation (Spearman, 1904) of 0.86 ($p < 0.001$) suggested that the content of experiences following traumatic grief as represented tapped by the item pool of the IES was similar across both types of events and patient versus nonpatient populations.

A factor analysis was conducted using a principal factors procedure with a varimax rotation to assess the item assignments on the intrusion and avoidance subscales. Two factors were extracted. For all items, the factor loading on the hypothesized factor was higher than it was on the other factor. This was taken as evidence of the coherence of the two subscale item sets. Reliability data were also reported in this cross-validation study. Coefficients of internal consistency were reported for both subscales for all three time points for the two groups both separately and combined. These coefficients ranged from 0.79 to 0.92.

An important component of this study was the thorough examination of the nature of the relationship between the intrusion and avoidance subscales. This was undertaken to compare inferences from either the total score that was used or the two subscale scores. The case was advanced that if correlations of 0.40–0.60 in magnitude for all six of the time by group conditions, then there would be little recommend maintaining separate subscale scoring. The empirical findings and the conceptual rationale both indicated that separate subscale scores be retained because even though in five of the six conditions the subscales were substantially correlated (ranging from 0.57 to 0.78), the patient sample at the pretherapy evaluation point produced a coefficient of only 0.15, a conspicuously different result. Consequently, these data suggested that using only a total score could well obscure important differences in symptomatic status across phenomenological domains.

Sundin and Horowitz (2002) recently summarized published research on the original IES's psychometric characteristics. They presented nonweighted averages across 18 studies of coefficient alpha. The result for the Intrusion subscale was 0.86 and 0.82 for the Avoidance subscale. This chapter also presents estimates of stability over time. They found that the longer the time interval between the test and the retest, the lower the coefficient. Nevertheless, none was below 0.51. In this analysis, Sundin and Horowitz did not consider the overall level of symptoms of the different samples, an issue raised by Zilberg et al. in their (1982) paper.

As well, this issue was not taken into account in the calculation of the nonweighted average correlation of the two subscales of 0.63 across the 18 studies. Evidence that this issue needs to be considered in the interpretation of results was highlighted in the review of factor analyses of the 15 items. In these 12 studies, it appeared that only seven supported the two factor structure, with three obtaining three factors (avoidance and numbing being separate), and two finding only a single factor. The latter is what would be expected from samples where the proportion of those having significant symptomatology is low, a matter described in more detail elsewhere (Weiss, 2004b). A summary of 18 studies presented the correlations between a variety of other measures of symptoms and intrusion and avoidance. Most were appropriate, though it appeared that divergent validity was an issue that required further study, since the correlations with general symptoms were larger than the average relationship of the two subscales. This is a finding that appears to not be limited to the IES, as much research concentrates on convergent validity and does not present analyses that would detract from the case for the adequacy of the measure or scale in question.

In a more detailed review of the literature on the IES (Weiss, 2004a), it was noted that the original IES has been the most widely used self-report measure of stress response or PTSD symptoms of reexperiencing and numbing and avoidance of any measure. As of May 2001, the PILOTS database reported its use in 1,147 studies. The next most frequent measure was the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). The next most frequent measure of PTSD symptoms was the Mississippi Scale for Combat-Related PTSD (Keane, Caddell, & Taylor, 1988) with fewer than half of the citations. A search of PsychINFO targeting only empirical studies revealed 515 citations. In that review it was not possible to determine if data using the original IES were collected or whether the study used the Impact of Event Scale-Revised (IES-R) (Weiss, 2004a; Weiss & Marmar, 1997) in any study appearing after 1996.

THE INITIAL PILOT WORK ON THE IES-R

Despite the usefulness of the original IES, complete assessment of the response to traumatic events required tracking of response in the domain of hyperarousal symptoms. Beginning with data from a longitudinal study of the response of emergency services personnel to traumatic events, including the Loma Prieta earthquake (e.g., Weiss, Marmar, Metzler, & Ronfeldt, 1995), a set of seven additional items, with six to tap the domain of hyperarousal, and one to parallel the DSM-III-R and now DSM-IV diagnostic criteria for PTSD were developed, piloted, and then used. These additional seven items were interspersed with the existing

seven intrusion and eight avoidance items of the original IES using a table of random numbers to establish placement. The IES-R comprises these 22 items, and was originally presented in the first edition of this reference work (Weiss & Marmar, 1997).

An important consideration in the construction of the revised IES was to maintain comparability with the original version of the measure as much as was possible. Consequently, the one week time frame to which the instructions refer in measuring symptomatic response was retained, as was the original scoring scheme of frequency – 0, 1, 3, and 5 for the responses of "Not at all," "Rarely," "Sometimes," and "Often." The only modification to the original items that was made was to change the item "I had trouble falling asleep or staying asleep" from its double-barreled status into two separate items. The first is simply "I had trouble staying asleep" and because of a somewhat higher correlation between it and the remaining intrusion items it was assigned it to represent the original item in the Intrusion subscale. The second item, "I had trouble falling asleep" was assigned to the new Hyperarousal subscale because of its somewhat higher correlation with the other hyperarousal items, its somewhat lower correlation with the intrusion items, and its more apparent link with hyperarousal than with intrusion. The six new items comprising the Hyperarousal subscale target the following domains: anger and irritability, jumpiness and exaggerated startle response, trouble concentrating, psychophysiologic arousal upon exposure to reminders, and hypervigilance. As mentioned earlier, the one new intrusion item taps the dissociative-like reexperiencing captured in true flashback-like experiences. The reader is referred to Weiss (2004a) for a summary of the internal consistency of the three subscale, all of which were strong, the pattern of item-total correlations, test–retest stability, which was also satisfactory, and communality of the interitem correlations.

On the basis of the experience with those data, and considerations of the insufficiency of frequency as a completely summarizing marker for self-report, the over-weighting of responses of "Sometimes" and "Often" in the scoring scheme, the IES-R molted into a measure with the following characteristics: (1) the directions were modified so that the respondent is not asked about the frequency of symptoms in the past 7 days but is instead asked to report the *degree of distress* of the symptom in the past 7 days; (2) the response format was modified to a 0–4 response format with equal intervals – 0 = Not at all, 1 = A little bit, 2 = Moderately, 3 = Quite a bit, 4 = Extremely – rather than the unequal intervals of the original scale; and (3) the subscale scoring was changed from the sum of the responses to the mean of the responses, allowing the user to immediately identify the degree of symptomatology merely by examining the subscale scores, since they are presented in the same metric as the item responses, something the original scale did not. These changes brought the IES-R in

parallel format to the SCL-90-R (Derogatis, 1994), allowing for direct comparison of endorsement of symptom levels across these two instruments.

ISSUES IN ALTERNATE VERSIONS IN A DIFFERENT LANGUAGE

One of the key tasks of any research project is to insure that the instruments used are reliable and valid measures of the phenomenon or construct under study. For a variety of reasons, many measures that have contributed to the growing cross-cultural literature in traumatic stress and PTSD were initially developed in English, the IES-R being no exception. Consequently, for use with samples whose native language is not English, a translation of the measure is required. Given this requirement, it is useful to review some of the issues involved in that process.

In a recent paper, Mallinckrodt and Wang (2004) present a thorough and thoughtful review of some of the most important issues, as well as making some recommendations. These authors cite Hambleton's work on the difference between literal translation and what is described as *adaptation* of items from one language to another. For example, the English phrase "go on automatic pilot" if translated literally into German, will not give the sense of engaging in behaviors without active deliberation that is only recognized after the fact. Thus, if only a literal translation were adopted, the reliability and validity would be compromised. Mallinckrodt and Wang (p. 369) present Hulin's view that "[a] pair of items from the original scale and its adapted version are said to be equivalent when two individuals with the same amount or level of the construct being measured have equal probabilities of making the same response to the different language versions of the same item."

Flaherty et al. (1988) suggested that there were five levels of equivalence that an adapted measure should possess in order to show that it has cross-cultural validity. The first, *content equivalence*, involves establishing that the content domain of items is relevant and appropriate for both cultures. The second, they describe as *semantic equivalence*, establishing that each item of the new measure communicates the meaning of its parallel item on the original scale. The third is more methodological: *technical equivalence* addresses the question of whether the data collection method (e.g., self-report) produces comparable results in each culture. The fourth, *criterion equivalence*, involves evidence of parallel comparisons to within-culture norms. The fifth and final equivalence is *conceptual*. This addresses whether the construct or phenomenon has the same meaning in each culture. The claim of ordered equivalence posits that subsequent levels of equivalence cannot be achieved in the absence of equivalence in all prior levels.

For at least the last three decades, back-translation has been the method most commonly used to adapt a measure developed in one language and culture to another (Brislin, 1970). The approach involves first translating the measure from the original to the new language, then having the translated measure retranslated back into the original language, and finally verifying the retranslated measure with the original measure. More recently (Hambleton, 1994), the International Test Commission has offered guidelines for the translation process. First, the team of translators should comprise individuals who are fluent in both the original and proposed language and in addition these individuals should have some familiarity or expertise with the construct being measured. If at all possible, the team members should be native members of the proposed culture. As well, they should have familiarity with the principles of scale development and item writing. An especially desirable characteristic would be the capacity to view constructs in light of what Gough (1966, 1990) termed *folk concepts*. Folk concepts are nontechnical deeply embedded dimensions of personality that ordinary members of a culture appreciate and understand intuitively, concepts such as dominance, ambition, or attractive.

As in other judgment tasks such as the determination of dynamic formulations (DeWitt, Kaltreider, Weiss, & Horowitz, 1983) or making diagnoses (Weiss, 2003), individuals doing the translation should work independently and not confer prior to completing the task. Mallinckrodt and Wang note that some have suggested that at this step if both the original and the translation are considered simultaneously, rather than using the original as the criterion, that the resulting translation will have a better chance of not floundering on problems involved with literal translation.

The next step is to have a completely separate team of translators (with the same attributes as the first) start with the translated product from the first team and to construct a back-translation into the original language. Obviously, they should be unfamiliar with the original measure. As before, the members should first work alone, and then confer to produce a final scale.

The final step is to have an expert panel compare the original and back-translated versions, with the same safeguards for independence as in the previous two steps. If the results are deemed to be equivalent then the process can terminate. If the results are not sufficiently equivalent, then the whole process should be repeated until an acceptable level of equivalence is obtained. Mallinckrodt and Wang also point out that the instructions must be viewed as an integral part of the process.

In their presentation, Mallinckrodt and Wang offer suggestions for the next phase of the process, one that they describe as quantitative verification using a *dual-language, split-half* approach (p. 370). This procedure closely resembles an approach offered by Norman (1965) for verification

of prediction models in regression approaches and is based on Mosier's methodology. This procedure comprises seven elements.

First, a large (exceeding 300 individuals) criterion sample that has provided data at the item level on the original measure is identified. They note that if construct validation measures are also available in this dataset that this is preferable.

Second, a bilingual sample is identified and checked for bilingual fluency. At least 30 individuals who identify with the target culture are advised.

Third, two forms of the measure are created such that for each form half of the items are derived from the original measure and the other half stem from the adapted measure with the item order such that items from one or the other source are all grouped together. This step also contains the proviso that for measures with a subscale structure, such as the IES-R, the two forms should contain items from all the subscales with as balanced a representation as possible. To counterbalance order effects of language, both forms in this step are used to generate two additional forms with the order of language reversed. Thus, if O1S1 = half of the original language items from subscale 1, O2S1 = remaining half of original language items from subscale 1, O1S2 = half of the original language items from subscale 2, O2S2 = remaining half of original language items from subscale 2, A1S1 = half of the adapted language items from subscale 1, A2S1 = remaining half of adapted language items from subscale 1, A1S2 = half of adapted language items from subscale 2, and A2S2 = remaining half of the adapted language items from subscale 2, the four versions of the measure will have the following formats: (1) combination of O1S1 and O1S2 followed by A1S1 combined with A1S2, (2) combination of A2S1 and A2S2 followed by O2S1 combined with O2S2, (3) combination of A1S1 and A1S2 followed by O1S1 combined with O1S2, and (4) combination of O2S1 and O2S2 followed by A2S1 combined with A2S2.

Fourth, each of the 30 members of the bilingual sample is randomly given one of the above four formats to complete, along with any other measures being utilized in the process (e.g., measures of convergent or divergent validity).

Fifth, if possible, data on test–retest stability are collected. Ideally, the interval between administrations should be the same for all 30 participants. Mallinckrodt and Wang suggest that the form for retest either be identical across administrations or describe a more complicated and time consuming optional approach. Here, all 30 participants complete two forms at the initial test session and the remaining two forms at the retest.

Sixth, the data are analyzed to produce the following results: (a) dual-language split-half reliability, (b) coefficient alpha internal consistency, (c) test–retest reliability, and (d) the analog of a multitrait multimethod matrix approach (Campbell & Fiske, 1959) for examining construct validity.

For the comparison of the two alpha coefficients, an *F*-test proposed for this purpose is recommended (Feldt, 1969). The approach recommended for the split-half reliability, test–retest reliability, and construct validity is one that has been employed elsewhere by the author (Weiss, 1979): the actual correlation coefficients become the raw data and comparison of these coefficients is accomplished using the *r* to *z* transformation.

The seventh and last element in the approach is practical only if a sample of appropriate native speakers large enough for the number of items to be examined is available (see Tinsley & Tinsley, 1987). The recommendation is to conduct a confirmatory factor analysis with the adapted version comparing its dimensionality to that found in the original version's criterion sample, or if the dimensionality is in question, to the several competing solutions. One of the vexing issues in this approach is that the assessment of goodness of fit of these models is considerably less clear than is commonly appreciated or acknowledged (Tomarken & Waller, 2003).

For the IES-R, the question of what model is most appropriate is not a straightforward decision. From the evidence so far, the structure of the scales may well be dependent on the nature of the sample being studied. The work of Zilberg et al. (1982) reported earlier clearly showed that the relationship between Intrusion and Avoidance varied as a function of level of distress and time elapsed since the event. In another set of data (Creamer, Bell, & Failla, 2003), the same phenomenon was noted: ". . . correlations among the subscales were higher in the community sample than in the treatment sample" (p. 1489) and this led to variations in the best-fitting models.

Indeed, it may well be the case not only for the IES-R, but for many if not most symptom measures of PTSD, the phasic nature of the symptom pattern, its longitudinal course, and the polythetic nature of the diagnosis make it less than clear what the structure of any measure ought to be. Because the time elapsed since the exposure to the traumatic event exercises such a significant impact on the symptom presentation. As well, given the reality that what Meehl (1995) termed the definitory criteria for PTSD are not known, the problem is more complex. Given that the field has accepted as defining of PTSD use of what are for disorders with known etiology (e.g., AIDS) merely evidentiary criteria (e.g., decreased CD-4 count, wasting, opportunistic infections, dementia), the nature of the structure of evidentiary characteristics has wholly different meanings and implications. Another example is Alzheimer's disease. The definitory characteristics are the plaques and tangles. Symptomatic manifestations, such as memory problems, acting out, and sunsetting, are not used to structure the nature of the disorder. Rather, the issue is commonness or uncommonness of course or symptom presentation. This viewpoint is presented in more depth elsewhere (Waller & Meehl, 1998) and cannot be

developed further here. It is sufficient to note, however, that the inference pattern of structure used for ability or intelligence, or even personality, probably does not well map onto situations where there are distinct subgroups whose clustering of symptoms is of key significance.

INTERNATIONAL VERSIONS OF THE IES-R

The efficiency and directness of the IES-R has led scholars in a variety of different countries and cultures to produce versions in non-English languages. A review of the literature revealed that the work accomplished for the international versions approached the recommendations of Mallinckrodt and Wang (2004) to varying degrees. It is, of course, an empirical question as to whether the detailed and extensive approach suggested by these authors would produce a more reliable or valid version than a more manageable approach. Published data are in the literature for formal translations as well as ad hoc translations in the context of an investigation of another question. Of the former, the following versions can be found (listed alphabetically): Chinese (Wu & Chan, 2003), French (Brunet, St-Hilaire, Jehel, & King, 2003), German (Maercker & Schuetzwohl, 1998), Japanese (Asukai et al., 2002), and Spanish (Baguena et al., 2001). A Bosnian version of the IES-R is described in a study that compares refugees to nonrefugee (Hunt & Gakenyi, 2005). Nonpublished (as of this writing) versions exist in Dutch (S. Bal, personal communication, September 23, 1998), Italian (Giannantonio, 2003), Norwegian (as cited in Kanagaratnam, Raundalen, & Asbjornsen, 2005), and Persian (Panaghi, Hakimshooshtary, Attari moghadam, & Ghorbani, 2005). There may well be other versions as well, as there has been informal communication with the author considering translations into Lithuanian, Portuguese, and Turkish.

Internal Consistency. The published results of the various international versions reveal reassuring consistency in the basic psychometric characteristic of internal consistency of the IES-R. In the Chinese version, for example, coefficient alpha was 0.89 for the Intrusion subscale, 0.85 for the Avoidance subscale, and 0.83 for Hyperarousal subscale. In the initial presentation of the French version, coefficient alpha was 0.86 for the Intrusion and Avoidance subscales, and 0.81 for the Hyperarousal subscale. The German version produced coefficient alpha of 0.87 for the Intrusion subscale, 0.78 for the Avoidance subscale, and 0.87 for the Hyperarousal subscale. The initial presentation of the Japanese version presented coefficients for four different samples: Intrusion – 0.91, 0.88, 0.89, and 0.91; Avoidance – 0.88, 0.81, 0.84, and 0.90; and Hyperarousal – 0.86, 0.80, 0.80, and 0.86. In the aggregate, as well individually, all the coefficients reveal considerable subscale homogeneity. These results are consistent with the

outcome of viable versions of the measure, and are what would be expected given the results of the English version (Weiss, 2004a).

Test–Retest Stability. The data regarding cross-time stability, an important characteristic of reliability from the perspective of reproducibility, were also consistent and encouraging. The initial publication of the Chinese version reported these data: $r = 0.74$ for Intrusion, $r = 0.52$ for Avoidance, and $r = 0.76$ for Hyperarousal. The French translation reported $r = 0.73$ for the Intrusion subscale, $r = 0.77$ for the Avoidance subscale, and $r = 0.71$ for the Hyperarousal subscale. The data collected for the Japanese version did not examine stability in as differentiated an approach as the other versions did. Instead of examining each subscale separately, only the total score of the three subscales was analyzed, and a Spearman (1904) rank order correlation was presented. In the sample of 114 participants, the data yielded $r_s = 0.86$. As for the findings with internal consistency, the stability data are what would be expected based on the findings from the original version.

Scale Intercorrelations. Zilberg et al. (1982) showed that the correlations of intrusion and avoidance in the original IES varied as a function of time elapsed since the traumatic event and level of symptomatology. Though the data from the French and Chinese translations (the Japanese translation did not present these data) could not address the correlations in this differentiated manner, they did, nonetheless, report the subscale correlations. In the French version, the correlation of Intrusion with Avoidance was $r = 0.62$ and with Hyperarousal was $r = 0.69$. The correlation of Avoidance with Hyperarousal was $r = 0.56$. The analogous data in the Chinese version were $r = 0.76$, $r = 0.83$, and $r = 0.75$. For the German version, in a same sample of former political prisoners, correlations of 0.61 (Intrusion and Avoidance), 0.85 (Intrusion and Hyperarousal), and 0.65 (Avoidance and Hyperarousal) were reported.

Scale Structure. In the initial reports of the international versions, analyses were presented that proceeded beyond the simple subscale intercorrelations. A summary of these is presented below. These results should, however, be viewed in light of the comments regarding evidentiary characteristics as compared to descriptive characteristics as described above.

The Chinese version data yielded a single strong factor that accounted for 45% of the variability in the item set. The data from the French version were subjected to a principal components analysis and with a varimax rotation. The results were not definitive. Both a two factor and a three factor solution were interpreted. The two factor solution comprised an avoidance factor and a combined intrusion–arousal factor. This was similar to the structure found in the Spanish version. The three factor solution replicated the three symptom criteria of PTSD: hyperarousal, avoidance, and intrusion. Item loadings (the correlation of the item with the score on the factor) for the set of 22 items were almost completely

coherent with each item loading most strongly on its own scale, though there were instances of low communality (the item did not go with any of the others in these data).

In contrast, the German version produced data that were most consistent with a four factor result. These were interpreted as factors measuring intrusion, avoidance, hyperarousal, and a separate numbing dimension. This is the same result as one reported for the Clinician-Administered PTSD Scale by King and his colleagues (1998).

The Japanese data used a kind of factor analysis termed "Varclus," marketed by the SAS Institute (1999), that attempts to find groups of variables that are as correlated as possible among themselves and as uncorrelated as possible with variables in other clusters. The key difference is that all variables start in a single cluster, and additional clusters are formed based on parameters set by the user. The paper reporting the Japanese reported neither the extraction method (principal components versus centroid) nor anything other than a forced three cluster solution. The results suggested that a model comprising three clusters of items fit those data best: an intrusion–hyperarousal cluster, an avoidance cluster, and a third cluster of numbing and sleep and cognitive distress. These clusters were not orthogonal. The correlation of cluster 1 and 2 was $r = 0.74$, 1 and 3 was $r = 0.73$, and 2 and 3 was $r = 0.62$. These correlations are of roughly the same magnitude as the regular subscales reported above.

The conclusion to draw from this set of analyses is that although the international versions of the IES-R show very similar basic psychometric properties in terms of internal consistency, stability, and subscale correlations, at the level of analysis of underlying dimensions, the picture is more complicated. Some of this variability is no doubt due to sample differences in size, homogeneity, and level of distress, but it well may be that differences in exposure to the trauma under study, differences in time elapsed since the exposure, and differences in comorbidity contribute to this pattern of results. Though the model that Mallinckrodt and Wang (2004) suggests that such variability may be a cause for concern, it is important to reiterate that at the level of structure, trait versus state issues may be more of an issue. More research about this topic would be welcome and hopefully clarifying.

A SAMPLING OF USE OF THE IES-R INTERNATIONAL VERSIONS

Keeping pace with the burgeoning PTSD literature, there is a growing literature using international versions of the IES-R. These studies use both English and other language versions, and cover a wide range of events, populations, age ranges, and research questions. The presentation below

is by no means exhaustive, and instead is merely a sampling, with no claims as to representativeness as to topic, wideness of usage, or other grouping or categorical variables. The order of presentation is by year of publication and within year alphabetically by first author's last name.

The German version of the IES-R was utilized in a study of individuals living in Germany who had had a life-threatening cardiac event and received an implanted cardioverter defibrillator (Baumert, Simon, Gündel, Schmitt, & Ladwig, 2004). The authors directed special attention to the relationships among the subscales as well as an examination of the concordance of scores on the Hyperarousal subscale with psychophysiological responses within the context of an acoustic startle reflex paradigm. The decision was made to use an outdated scoring algorithm (0, 1, 3, 5), thus rendering problematic the generalizability of the conclusions the authors drew. An attractive feature of the study was the use of a receiver operating characteristic (ROC) (see Kraemer, 1992) approach to classification regarding hyperarousal. The data revealed that the internal consistency of the Intrusion and Avoidance subscales was above 0.80, whereas the coefficient was 0.66 for the Hyperarousal subscale. In these data, the ability of scores on the Hyperarousal subscale to distinguish those who had strong psychophysiological responses to the acoustic startle paradigm from those who had weak responses was disappointing. In fact, the operation of the Hyperarousal subscale in this sample was less strong than either Intrusion or Avoidance. The authors acknowledge the possibility that because this sample attained lower scores on the subscales than more standard traumatic events typically produce, this may have affected the characteristics of the measure in unknown ways.

A second study in Germany using the German version examined similarities and differences among victims of a recent trauma, patients with PTSD, and healthy controls regarding deployment of attention, heart rate responses and self-ratings to trauma relevant pictures (Elsesser, Sartory, & Tackenberg, 2004). This was a complex study employing measures of heart rate and electromyogram assessment of eye blink response to acoustic startle. A dot probe task, and color picture task included trauma relevant, generically aversive and emotionally neutral stimuli to help shed light on the hyperarousal aspect of PTSD, and, in this study, Acute Stress Disorder. The total sample comprised 86 individuals divided as 37 with recent exposure, 18 with PTSD that averaged a bit more than 2 years time elapsed, and 31 healthy controls. There were a number of findings, including a tendency to show heart rate acceleration to trauma-related material in both exposed groups. Of special interest regarding the IES-R, PTSD patients obtained higher scores than the group who had experienced recent trauma. As well, the investigators demonstrated that those with more intrusions showed a bias away from trauma-relevant presentations, suggesting that the Intrusion subscale

can capture differences that are manifested in standard experimental behavioral paradigms.

Hunt and Evans (2004) were interested in the constructs of emotional intelligence and the phenomenon of monitoring versus blunting with respect of symptoms after traumatic exposures in a study conducted in the United Kingdom. Individuals who obtain high scores on monitoring are active processors of traumatic events and are open to information, whereas individuals who tend to avoid information are described as employing a blunting approach. Hunt and Evans used the standard IES-R with the older scoring algorithm and gathered data on a sample of 442 individuals of whom 233 were female. Of the 442, 298 reported exposure to a traumatic event, and it is for this subsample that the IES-R findings apply. Statistically significant higher scores on all three subscales were obtained by the women in the study. The effect sizes (Rosenthal, 1994) for Intrusion, Avoidance, and Hyperarousal were 0.30, 0.36, and 0.43, respectively. There was no relationship between a total score across the three subscales (a variable some researchers choose to compute) and the measure of monitoring ($r = -0.03$), but there was a significant but small effect for blunting ($r = 0.14$, $p < 0.005$). The authors also hypothesized that those with higher emotional intelligence scores would report less symptomatic distress. These findings were stronger than those for monitoring and blunting. The correlation was –0.31 for the Intrusion subscale, –0.29 for the Avoidance subscale, and –0.32 for the Hyperarousal subscale.

The Sarin gas attack in the Tokyo subway in 1995 provided the opportunity to examine symptomatic status at 5 years follow-up for 34 exposed individuals (Ohtani et al., 2004). The Japanese version of the IES-R was utilized in this study, as was the Clinician-Administered PTSD Scale (Blake et al., 1995; Weathers et al., 2001). The investigation examined a potential pool of 565 exposed who were treated in the emergency room at the time of exposure. Responses totaled 170, with 64 agreeing to consider participating. Of these, 34 provided full data. The authors presented item-level data for the IES-R, providing an unusual opportunity to look at specific symptoms. The authors reported that more individuals (76%) endorsed some level of hypervigilance ("I felt watchful and on guard") than any other item. The item least endorsed (12%) was from the Avoidance subscale – "I felt as if it hadn't happened or wasn't real." The two sleep disturbance items were tied for the percentage of survivors (9%) who indicated extreme distress in the last week. Looked at from the perspective of the subscales, there was little differentiation in the subscales regarding the average percentage of individuals who reported no symptomatology when pooled across all the items within each subscale: Intrusion, 57%; Avoidance, 59%; and Hyperarousal, 56%. Finally, there was evidence of construct validity in the relationship between the IES-R and the Clinician-Administered PTSD scale.

This same sample provided a subsample of individuals whose auditory P300 event-related potentials were related to brain morphology, specifically anterior cingulate gray matter volume (Araki et al., 2005). In this study, the IES-R was part of the battery that was compared between those who carried a diagnosis of PTSD (n = 8) and those who did not (n = 13). Using the Japanese version, and using only total score, the authors reported a statistically significant difference (t = 2.68, $p < 0.017$) with an effect size of 1.3, large effect using the metric recommended by Cohen (1988). The key finding of this research was that those with PTSD showed significantly lower P300 amplitudes in response to an oddball task; these scores were not, however, associated with scores on the IES-R. As well, there was some suggestion that the degree of lowered P300 amplitude was related to reduced volume in the brain area examined – the anterior cingulate. This latter finding, however, though intriguing, cannot at this point be thought of as established.

The English version of the IES-R was utilized in an investigation among Jerusalem residents seen in an emergency room following exposure to a traumatic incident (Bachar, Hadar, & Shalev, 2005). The question under investigation in this project was whether a measure of narcissistic traits and vulnerability would show some association with the development of PTSD. The authors showed that those who were more symptomatic at 1 month, and 4 months, had significantly higher levels of narcissistic vulnerability at baseline were. Because the focus of the report was on the scale to measure narcissistic vulnerability, actual coefficients between it and the subscales of the IES-R were not specifically reported. Instead, these variables and the Beck Depression Inventory were reported to correlate with the vulnerability measure in the range of 0.24–0.39, across both time points.

The German version of the IES-R was utilized in a study at the Medical University of Vienna, in Austria (Bunzel, Laederach-Hofmann, Wieselthaler, Roethy, & Drees, 2005). The sample comprised long-term survivors of heart transplant with a period of uncertainty before the transplant that life was sustained by a circulatory assist device and their partners who were assessed retrospectively. These authors chose to use a cut-off score from the total score to estimate a diagnosis of PTSD. They found that none of the patients but 23% of the partners achieved a score high enough to be termed having PTSD. Concurrently, therefore, the partners scored significantly higher than the transplant patients on all three subscales. The effect size was 0.82 for Intrusion, 0.56 for Avoidance, and 0.82 also for Hyperarousal. In this study the IES-R was able to detect effects between differentially exposed groups, though it might have been thought that those who were ill would have found the incident more traumatic than the helpless partner.

The French version was part of the assessment package in a study of adolescents in Toulouse, France, following the explosion of a chemical

factory so strong that it produced a tremor measuring 3.4 on the Richter scale (Godeau et al., 2005). These authors confined their use of the IES-R to those participants 15 and 17 years old, though the directly exposed sample of 577 comprised 11- and 13-year olds as well. The study focused on what the authors termed "symptoms consistent with PTSD" (SCW-PTSD), thus avoiding the thorny issue of diagnosis via self-report, but this was just another way to described symptomatic elevations. They compiled a control sample of 900 nondirectly exposed children and conducted the survey 9 months after the event. Among many results, they found that almost 30% of 15- and 17-year olds who were directly exposed had salient symptoms as compared to approximately 5% in the nondirectly exposed. Within those directly exposed, older children had higher symptoms levels and girls had twice the level of symptoms as boys. There were other results related to being injured, having severe damage at school or at home, and having a cumulative impact of multiple consequences, the participant and a family member both injured, compared to one or no injuries.

Hong Kong was the locale for a study utilizing the Chinese version the IES-R and severe acute respiratory syndrome (SARS) was the topic at issue (Ho, Kwong-Lo, Mak, & Wong, 2005). Only one of two samples studied included data about the IES-R. This group comprised 97 staff members who had been infected with SARS, a response rate of 30% of those approached. The study was focused on a measure of fear of SARS with aspects of Infection, Insecurity, and Instability as well as a total score. There were robust correlations between Intrusion, Avoidance, and Hyperarousal and all four fear variables, ranging from a low of 0.23 between Avoidance and Infection, and a high of 0.66 between Intrusion and total score. On average, the highest correlations for the four Fear variables were with Intrusion and the lowest was for Avoidance. Elevations on the subscales ranged from 1.24 to 1.57, which were higher than in the initial validation sample (Wu & Chan, 2003).

The Bosnian version was employed in a study of refugees and nonrefugees (Hunt & Gakenyi, 2005). The sample of refugees showed higher levels on symptoms than nonrefugees, with those who were older showing higher scores. Unlike in many other studies, in this sample of 69 refugees and 121 nonrefugees, there was no difference in scores as a function of gender. The authors in this project also chose to adopt a cut-off score, though they used a nonstandard metric so the exact value is not particularly illuminating. Nevertheless, they found that 77% of the refugees as opposed to 45% of nonrefugees exceeded their categorical designation. Even when age and traumatic experience were controlled, refugee status still produced higher scores than nonrefugee status. This study also examined the effect of personality variables. There was a significant correlation between a measure of Harmavoidance and Intrusion (r = 0.17) and Hyperarousal (r = 0.28), but not with Avoidance.

Some cross-cultural studies examine a group of individuals originally from one culture in another culture. This situation characterizes a study of

20 adult former Tamil rebels from Sri Lanka being studied in Norway (Kanagaratnam et al., 2005) using the Norwegian version of the IES-R but with an outdated scoring algorithm. The focus of the researchers was on symptoms and ideological commitment. The sample was approximately 25 years old and comprised only two women. They had been in Norway an average of 69 months. In this study, the relationship symptom level and other factors were assessed via a rank order correlation. Neither age, nor length of training, nor length of exposure, nor time elapsed, nor time in Norway was associated with level of distress, which did show considerable variability, so restriction of range is not a strong explanatory factor. As the authors anticipated, weaker commitment to the Tamil cause was associated with significantly lower levels of Avoidance and Hyperarousal, but not Intrusion.

SUMMARY AND CONCLUSIONS

The IES-R has generated a number of formal international versions, several informal versions that have appeared in the context of a topically oriented peer-reviewed publication, and a number of unpublished international versions. At the level of basic psychometric properties, the published data suggest impressive concordance in terms of internal consistency, test–retest stability, and subscale correlations even though the methods used have not employed all aspects of a comprehensive and exhaustive approach that is admittedly challenging and expensive to undertake. The relationships of the subscales to each other also appear to be relatively similar across versions. Findings regarding the underlying dimensions of the item pool are less coherent. One reason is the decision of the researcher to seek dimensions that are independent or correlated, since these dictate different analytic strategies. A second has to do with whether the emphasis is empirical or conceptual. A third has to do with individual characteristics of samples, traumatic events, prevalence of problems, time elapsed, and similar issues. More extensive data will help clarify this state of affairs.

There is a broad and growing literature using the IES-R in both the original English version in a variety of nationalities and cultures, as well as international versions in a variety of different languages in a set of diverse and different cultures. The picture that emerges is of a robust and interesting set of findings that will undoubtedly continue to grow and expand. Cautions in interpreting the literature include the continuing use of nonrecommended metrics for scoring item responses (e.g., 0, 1, 3, 5), forming scale scores by summing items rather than taking the mean, thereby making results not comparable across studies, as well as other variations including changes in directions, time-frame, incident description, and the other items enumerated elsewhere (Weiss, 2004a).

REFERENCES

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

Araki, T., Kasai, K., Yamasue, H., Kato, N., Kudo, N., Ohtani, T., et al. (2005). Association between lower P300 amplitude and smaller anterior cingulate cortex volume in patients with posttraumatic stress disorder: A study of victims of Tokyo subway sarin attack. *Neuroimage, 25*, 43–50.

Asukai, N., Kato, H., Kawamura, N., Kim, Y., Yamamato, K., Kishimoto, J., et al. (2002). Reliability and validity of the Japanese-language version of the Impact of Event Scale-revised (IES-R-J): Four studies of different traumatic events. *Journal of Nervous and Mental Disease, 190*, 175–182.

Bachar, E., Hadar, H., & Shalev, A. Y. (2005). Narcissistic vulnerability and the development of PTSD: A prospective study. *Journal of Nervous and Mental Disease, 193*, 762–765.

Baguena, M. J., Villarroya, E., Belena, A., Amelia, D., Roldan, C., & Reig, R. (2001). Psychometric properties of the Spanish version of the Impact of Event Scale-Revised (IES-R)/Propiedades psicometricas de la version espanola de la Escala Revisada de Impacto del Estressor (EIE-R). *Analisis y Modificacion de Conducta, 27*, 581–604.

Baumert, J., Simon, H., Gündel, H., Schmitt, C., & Ladwig, K.-H. (2004). The Impact of Event Scale-Revised: Evaluation of the subscales and correlations to psychophysiological startle response patterns in survivors of a life-threatening cardiac event: An analysis of 129 patients with an implanted cardioverter defibrillator. *Journal of Affective Disorders, 82*, 29–41.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

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

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 185–216.

Brunet, A., St-Hilaire, A., Jehel, L., & King, S. (2003). Validation of a French version of the Impact of Event Scale-Revised. *Canadian Journal of Psychiatry, 48*, 56–61.

Bunzel, B., Laederach-Hofmann, K., Wieselthaler, G. M., Roethy, W., & Drees, G. (2005). Posttraumatic stress disorder after implantation of a mechanical assist device followed by heart transplantation: Evaluation of patients and partners. *Transplantation Proceedings, 37*, 1365–1368.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Creamer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the Impact of Event Scale-Revised. *Behaviour Research & Therapy, 41*, 1489–1496.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Deebiec, J., & Ledoux, J. E. (2006). Noradrenergic signaling in the amygdala contributes to the reconsolidation of fear memory: Treatment implications for PTSD. *Annals of the New York Academy of Sciences, 1071*, 521–524.

Derogatis, L. R. (1994). *Symptom Checklist-90-R*. Minneapolis, MN: National Computer Systems.

DeWitt, K. N., Kaltreider, N. B., Weiss, D. S., & Horowitz, M. J. (1983). Judging change in psychotherapy. Reliability of clinical formulations. *Archives of General Psychiatry, 40*, 1121–1128.

Elsesser, K., Sartory, G., & Tackenberg, A. (2004). Attention, heart rate, and startle response during exposure to trauma-relevant pictures: A comparison of recent trauma victims and patients with posttraumatic stress disorder. *Journal of Abnormal Psychology, 113*, 289–301.

Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder–Richardson coefficient twenty is the same for two tests. *Psychometrika, 34*, 363–373.

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

Giannantoio, M. (a cura di) (2003). *Psicotraumatolgia e psicologia dell'emergenzu*. Salerno, Ecomind.

Godeau, E., Vignes, C., Navarro, F., Iachan, R., Ross, J., Pasquier, C., et al. (2005). Effects of a large-scale industrial disaster on rates of symptoms consistent with posttraumatic stress disorders among schoolchildren in toulouse. *Archives of Pediatrics & Adolescent Medicine, 159*, 579–584.

Gough, H. G. (1966). A cross-cultural analysis of the CPI Femininity scale. *Journal of Consulting Psychology, 30*(2), 136–141.

Gough, H. G. (1990). The California psychological inventory. In C. E. Watkins, Jr. & V. L. Campbell (Eds.), *Testing in counseling practice. Vocational psychology* (pp. 37–62). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10*, 229–244.

Ho, S. M. Y., Kwong-Lo, R. S. Y., Mak, C. W. Y., & Wong, J. S. (2005). Fear of severe acute respiratory syndrome (SARS) among health care workers. *Journal of Consulting and Clinical Psychology, 73*, 344–349.

Horowitz, M. J. (1975). Intrusive and repetitive thoughts after experimental stress. A summary. *Archives of General Psychiatry, 32*, 1457–1463.

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

Horowitz, M. J., & Kaltreider, N. B. (1977). The response to stress. *Comprehensive Therapy, 3*, 38–40.

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

Hunt, N., & Evans, D. (2004). Predicting traumatic stress using emotional intelligence. *Behaviour Research & Therapy, 42*, 791–798.

Hunt, N., & Gakenyi, M. (2005). Comparing refugees and nonrefugees: The Bosnian experience. *Journal of Anxiety Disorders, 19*, 717–723.

Kanagaratnam, P., Raundalen, M., & Asbjornsen, A. E. (2005). Ideological commitment and posttraumatic stress in former Tamil child soldiers. *Scandinavian Journal of Psychology, 46*, 511–520.

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

Keane, T. M., Kolb, L. C., Kaloupek, D. G., Orr, S. P., Blanchard, E. B., Thomas, R. G., et al. (1998). Utility of psychophysiology measurement in the diagnosis of posttraumatic stress disorder: Results from a department of Veteran's Affairs cooperative study. *Journal of Consulting and Clinical Psychology, 66*, 914–923.

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

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

Maercker, A., & Schuetzwohl, M. (1998). Assessment of post-traumatic stress reactions: The Impact of Event Scale-Revised (IES-R)/Erfassung von psychischen Belastungsfolgen: Die Impact of Event Skala-revidierte Version (IES-R). *Diagnostica, 44*, 130–141.

Mallinckrodt, B., & Wang, C.-C. (2004). Quantitative methods for verifying semantic equivalence of translated research instruments: A Chinese version of the experiences in close relationships scale. *Journal of Counseling Psychology, 51*, 368–379.

Meehl, P. E. (1995). Bootstraps taxometrics. Solving the classification problem in psychopathology. *American Psychologist, 50*, 266–275.

Norman, W. T. (1965). Double-split cross-validation: An extension of Mosier's design, two undesirable alternatives, and some enigmatic results. *Journal of Applied Psychology, 49*(5), 348–357.

Ohtani, T., Iwanami, A., Kasai, K., Yamasue, H., Kato, T., Sasaki, T., et al. (2004). Post-traumatic stress disorder symptoms in victims of Tokyo subway attack: A 5-year follow-up study. *Psychiatry and Clinical Neurosciences, 58*, 624–629.

Panaghi, L., Hakimshooshtary, M., Attari moghadam, J., & Ghorbani, B. (2005). Validation of the Persian version of Impact of Event Scale-Revised. Tehran: Iran University of Medical Sciences.

Pitman, R. K., Sanders, K. M., Zusman, R. M., Healy, A. R., Cheema, F., Lasko, N. B., et al. (2002). Pilot study of secondary prevention of posttraumatic stress disorder with propranolol. *Biological Psychiatry, 51*, 189–192.

Rosenthal, R. (1994). Parametric measures of effect size. In H. M. Cooper & L. V. Hedges (Eds.), *The Handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

SAS Institute Inc. (1999). *SAS/STAT user's guide*. Cary, NC: Author.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 14*, 72–101.

Sundin, E. C., & Horowitz, M. J. (2002). Impact of Event Scale: psychometric properties. *British Journal of Psychiatry, 180*, 205–209.

Tinsley, H. E., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research [Special issue: Quantitative foundations of counseling psychology research]. *Journal of Counseling Psychology, 34*, 414–424.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "Well Fitting" models. *Journal of Abnormal Psychology, 112*, 578–598.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage Publications.

Weathers, F. W., Keane, T. M., & Davidson, J. R. (2001). Clinician-administered PTSD scale: A review of the first ten years of research. *Depress Anxiety, 13*, 132–156.

Weiss, D. S. (1979). The effects of systematic variations in information on judges' descriptions of personality. *Journal of Personality & Social Psychology, 37*, 2121–2136.

Weiss, D. S. (2003). Research methodology and statistics. In A. Tasman, J. A. Lieberman, & J. Kay (Eds.), *Psychiatry* (2nd ed., Vol. 1, pp. 491–506). London: John Wiley & Sons.

Weiss, D. S. (2004a). The Impact of Event Scale-Revised. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD: A practitioner's handbook* (2nd ed., pp. 168–189). New York: Guilford Press.

Weiss, D. S. (2004b). Structured clinical interview techniques for PTSD. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD: A practitioner's handbook* (2nd ed., pp. 103–121). New York: Guilford Press.

Weiss, D. S., Marmar, C. R., Metzler, T. J., & Ronfeldt, H. M. (1995). Predicting symptomatic distress in emergency services personnel. *Journal of Consulting & Clinical Psychology, 63*, 361–368.

Weiss, D. S., & Marmar, C. R. (1997). The Impact of Event Scale – Revised. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD: A practitioner's handbook* (pp. 399–411). New York: Guilford Press.

World Health Organization. (1978). *Clinical modification of the World Health Organization's Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death* (9th revised ed.). Geneva: Author.

Wu, K. K., & Chan, K. S. (2003). The development of the Chinese version of Impact of Event Scale – Revised (CIES-R). *Social Psychiatry & Psychiatric Epidemiology, 38*, 94–98.

Yehuda, R. (2006). Advances in understanding neuroendocrine alterations in PTSD and their therapeutic implications. *Annals of the New York Academy of Sciences, 1071*, 137–166.

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