The Impact of Event Scale-Revised: Psychometric properties in a sample of motor vehicle accident survivors

J. Gayle Beck a,*, DeMond M. Grant a, Jennifer P. Read a, Joshua D. Clapp a, Scott F. Coffey b, Luana M. Miller a, Sarah A. Palyo a

a Department of Psychology, University at Buffalo-SUNY, 230 Park Hall, Buffalo, NY 14260, United States
b Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, MS, United States

Received 26 October 2006; received in revised form 15 February 2007; accepted 21 February 2007

Abstract

This study examined the factor structure, internal consistency, concurrent validity, and discriminative validity of the Impact of Event Scale-Revised (IES-R, [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 (pp. 399–411). New York: Guilford Press]) in a sample of 182 individuals who had experienced a serious motor vehicle accident. Results supported the three-factor structure of the IES-R, Intrusion, Avoidance, and Hyperarousal, with adequate internal consistency noted for each subscale. Support was obtained for the concurrent and discriminative validity, as well as the absence of social desirability effects. Although some differences were noted between the IES-R Avoidance subscale and diagnostically based measures of this cluster of symptoms, these differences do not necessarily signify measurement problems with the IES-R. The IES-R seems to be a solid measure of post-trauma phenomena that can augment related assessment approaches in clinical and research settings.

# 2007 Elsevier Ltd. All rights reserved.

Keywords: IES-R; Post-traumatic stress disorder; Trauma

1. Introduction

The Impact of Event Scale (IES, Horowitz, Wilner, & Alvarez, 1979) is one of the most widely used self-report measures within the trauma literature (e.g., Joseph, 2000; Weiss & Marmar, 1997). The scale, published just before release of the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III, American Psychiatric Association, 1980) is grounded in Horowitz’s model of emotional processing following a trauma (Horowitz, 1976). According to this model, until traumatic experiences are psychologically assimilated, the individual will alternate between the experience of intrusive thoughts and feelings in one moment and avoidance strategies in the next. Following this model, the IES was constructed with two subscales, one tapping intrusions (e.g., repeated thoughts about the trauma) and the other tapping avoidance (e.g., repeated thoughts about the trauma) and the other tapping avoidance (e.g., effortful avoidance of situations that serve as reminders of the trauma). Shortly after the IES was published, post-traumatic stress disorder (PTSD) was introduced into the DSM-III. Importantly, the DSM-III conceptualization of PTSD diverged somewhat from Horowitz’s information processing model, and included three symptom clusters. In addition to a cluster of intrusion-based symptoms and a cluster of symptoms...
involving reduced involvement with the external world (i.e., avoidance) that the IES assessed, the DSM also delineated a third cluster of symptoms, physiological hyperarousal. This tripartite model of PTSD has persisted over the almost three decades since the DSM-III was first published. Despite other diagnostic revisions, subsequent editions of the DSM have continued to view PTSD as being comprised of these three distinct symptom clusters.

Despite the incomplete match between the symptoms measured by the IES and those viewed as defining diagnostic criteria for PTSD, the IES has become an important measurement tool within the trauma literature. It is used with both adult and child trauma survivors and has been translated into several other languages (Joseph, 2000; Sudin & Horowitz, 2003). However, some believe that the IES has a number of conceptual and measurement problems. First, as highlighted by Larsson (2000), the scale may assess general negative affectivity rather than any trauma-specific phenomena. As the scale was originally constructed to measure “the current degree of subjective impact experienced as a result of a specific event” (Horowitz et al., 1979, p. 209), the intrusion and avoidance subscales were not designed to represent independent constructs that would be distinct and unique from general distress. However, it is unclear to what extent these subscales assess trauma-specific phenomena. The lack of assessment of hyperarousal within the IES contributes further to uncertainty about how the measure corresponds to the tripartite model of PTSD (defined within the DSM). As well, some concern has been raised about relatively high intercorrelations between the Intrusion and Avoidance subscales. Within Horowitz’s and other models of PTSD (e.g., Brewin, Dalgleish, & Joseph, 1996; Foa, Steketee, & Rothbaum, 1989), a functional relationship is proposed between intrusion and avoidance, whereby avoidance helps the individual to regulate negative affect that is generated by intrusive reminders of the traumatic event. Given the hypothesized interrelationship between intrusion and avoidance, moderate to strong correlations between these two subscales would be expected.

Additionally, some authors have questioned the underlying factor structure of the IES. For example, Shevlin, Hunt, and Robbins (2000) report although that a model containing one general distress factor fits the structure of the IES to some extent, a model containing two factors (intrusion and avoidance) with cross-loading noted for several items appeared to be a better fit. Andrews, Shevlin, Troop, and Joseph (2004) argued that the optimal structure of the IES consisted of four first-order factors (intrusion, avoidance, numbing, and sleep), with one second-order factor which assessed general distress. Although this dialogue parallels debate concerning the structure of PTSD symptoms more generally (e.g., Asmundson et al., 2000; DuHamel et al., 2004), the factor structure of the IES itself has been questioned, separate from this more general discussion.

Finally and most germane for this paper, despite being widely used within the trauma literature, the IES has been criticized as falling short as a measure of PTSD as defined within the DSM (e.g., Joseph, 2000). The scale contains items which do not reflect DSM symptomatology and as noted previously, the hyperarousal symptom cluster is not represented. In an effort to make the IES more reflective of the tripartite symptom criteria outlined by DSM, the IES-Revised was developed (IES-R, Weiss & Marmar, 1997). In its final form, the IES-R contains eight intrusion and eight avoidance items, derived from the original IES, and adds to this six items assessing hyperarousal. The authors of the IES-R intended for the scale to be comparable with the original scale and so, made only minor changes to the intrusion and avoidance subscales. The aim of this revision was to improve the utility of the IES and its applicability to the DSM symptomatology for PTSD.

Although the IES-R was first introduced in 1997, it has generated surprisingly little psychometric evaluation. To date, it has been translated into Chinese (Wu & Chan, 2003), Japanese (Asukai et al., 2002), French (Brunet, St-Hilaire, Jehel, & Kind, 2003), and Spanish (Baguena et al., 2001). Each of these efforts has examined the associated factor structure of the IES-R, with results ranging from good support for the three scale structure in a female sample that had recently experienced a natural disaster (Brunet et al., 2003), to observation of a single factor structure among a non-traumatized sample attending a medical emergency service (Wu & Chan, 2003). In an effort to examine the psychometric properties of the IES-R among Vietnam veterans seeking treatment for PTSD as well as a community sample of veterans with similar demographic characteristics, Creamer, Bell, and Failla (2003) conducted a confirmatory factor analysis to assess the extent to which the IES-R items corresponded to an intrusion, an avoidance, and a hyperarousal solution. Results showed poor model fit. An exploratory factor analysis suggested either a single factor or two factors, assessing intrusion/hyperarousal versus avoidance, with no clear statistical advantage for either model. It is difficult to reconcile these findings with those of Brunet et al. (2003) who also involved a sample with
well-documented trauma exposure. Possibly, the fact that the sample used in Creamer et al. (2003) study had experienced combat-related trauma thirty years prior to assessment influenced these results. An additional factor analytic effort by Baumert, Simon, Günßel, Schmitt, and Ladwig (2004) did not distinguish between the goodness of fit of a three-factor model and a four-factor model (assessing intrusion, avoidance, hyperarousal, and numbing). Clearly, additional work is needed on this topic.

Owing to the relative lack of psychometric data on the IES-R, the discrepant findings regarding the factor structure of this measure, and concern about how the revised scale performs relative to the DSM criteria for PTSD, the current investigation had three goals. First, we examined the extent that the factor structure of the IES-R approximated the tripartite model of DSM in a sample that had experienced a traumatic motor vehicle accident (MVA). Owing to mixed findings in previous studies, it is important to examine this issue in a sample of individuals who had experienced the same trauma and who satisfy current diagnostic definitions of a traumatic experience, in order to clarify interpretation of any obtained factor structure. Survivors of a serious MVA were thought to be an appropriate sample for study given that MVAs are among the leading causes of PTSD in the United States (e.g., Norris, 1992). In this report, participants were assessed using a well-regarded diagnostic interview to assess PTSD, a methodological improvement over previous psychometric evaluations of the IES-R. We hypothesized that the three-factor structure of the IES-R would be confirmed in this sample, based on methodological similarities with Brunet et al. (2003) and with the samples involved during scale development (Weiss & Marmar, 1997).

The second goal of this report was to examine related psychometric features of the obtained factor structure, specifically internal consistency, concurrent validity using measures of PTSD symptomatology, anxiety, and depression, and the potential influence of social desirability. We hypothesized that the three subscales of the IES-R would show good concurrent validity with related measures of PTSD symptomatology derived from the DSM. In particular, we sought to examine cross-measure agreement within specific symptom clusters. We also hypothesized that the IES-R subscales would show moderate to high correlations with validated measures of anxiety and depression, given that PTSD symptoms have been shown to be associated with increases in other emotional disorders (e.g., Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). These analyses were repeated, controlling for litigation status, in order to determine whether the observed associations were partially influenced by the presence of legal action. The third goal of this study was to examine the ability of the IES-R to differentiate between individuals with and without diagnosable PTSD. In this effort, we examined the relative contributions of each subscale. Given that this was an exploratory analysis, no hypotheses were put forward for this aim.

2. Method

2.1. Participants

The sample included 182 individuals who sought assessment and possible treatment for mental health problems following their MVA. Participants were recruited from specialists in pain, rehabilitation, and internal medicine, as well as via public service announcements. Individuals qualified for assessment if they had experienced a MVA involving actual or threatened death or serious injury and their emotional response included intense fear, helplessness, horror, or the perception that they would die (Criterion A, American Psychiatric Association, 2000). These features were evaluated during initial phone contact with the project and using the MVA Interview (see below). Individuals involved in accidents that did not satisfy Criterion A were not evaluated. The sample included 142 women (78%) and 40 men (22%) and ranged in age from 18 to 79 years (mean 42.9, S.D. 12.87). One hundred and fifty-one participants (83%) were Caucasian, 24 (13%) were African-American, 3 (2%) were Hispanic, and 1 (5%) was Asian. The majority of patients (n = 133, 73%) reported on-going pain complaints from injuries sustained during the MVA. In these cases, pain caused significant lifestyle limitations (e.g., inability to work), impairment (e.g., use of prescription pain medications at least 3 days/week), or significant distress (e.g., continued health care utilization for pain). Average elapsed time after the MVA was 39.9 months (S.D. 83.4, Mdn = 11 months), with a range from 1 to 600 months. Thus, although most MVAs had occurred relatively recently, there was some variability in our sample in the time elapsed since the traumatic event. The majority of the sample (n = 129, 71%) was engaged in MVA-related litigation. Data from individuals presenting with neurological impairment, substance dependence and abuse in the 6 months preceding the assessment, psychotic symptoms, or acute suicidality were excluded. All participants provided informed consent prior to participation.
2.2. Measures

2.2.1. The IES-R

The Impact of Event Scale-Revised (Weiss & Marmar, 1997) is a 22-item scale which is rated on a 0 (not at all) to 4 (extremely) scale with respect to how distressing each item has been during the past week. Scale scores are formed for the three subscales, which reflect intrusion (eight items), avoidance (eight items), and hyperarousal (six items), and show a high degree of intercorrelation (r’s = .52–.87; Creamer et al., 2003). High levels of internal consistency have been previously reported (Intrusion: Cronbach’s alpha = .87–.94; Avoidance: Cronbach’s alpha = .84–.87; Hyperarousal: Cronbach’s alpha = .79–.91; Creamer et al., 2003; Weiss & Marmar, 1997). Test–retest reliability, collected across a 6-month interval, ranged from .89 to .94 (Weiss & Marmar, 1997). Similar internal consistency and test–retest values have been reported with a Japanese translation of the IES-R (Asukai et al., 2002).

2.2.2. Other PTSD measures

The MVA Interview was administered (Blanchard & Hickling, 1997), which includes questions about the individual’s emotional response to the accident (feelings of fear, helplessness, danger, and perceptions that they might die) to determine whether the MVA qualified as a traumatic event. Each of these responses to the MVA was rated on a 0–100 Likert Scale, where 0 = “not at all” and 100 = “extreme,” with a score of 50 or higher on ratings of fear, helplessness, or horror indicating that the MVA was experienced as traumatic.

PTSD symptomatology was assessed with both clinician and self-report measures. The Clinician-Administered PTSD Scale (CAPS, Blake et al., 1990), a structured interview that assesses PTSD symptoms identified in the current DSM-IV, was used as the diagnostic tool. The CAPS includes standardized questions to determine frequency and intensity. Symptoms were assessed in the preceding month, using a 5-point Likert Scale (e.g., 0 indicates “the symptom does not occur or does not cause distress” and 4 indicates “the symptom occurs nearly every day or causes extreme distress and discomfort”). The total severity score for the CAPS (CAPS-Total) is computed by summing the frequency and intensity ratings for each symptom (range 0–136). Probes were added to the interview to determine whether each PTSD symptom was attributable to pain (e.g., if a patient reported difficulty sleeping, the clinician assessed whether this symptom was due to pain. If so, the symptom was not scored on the CAPS).

The CAPS was administered by trained clinicians who were advanced psychology doctoral students. All interviews were videotaped and 35% (n = 63) were randomly selected and reviewed by an independent clinician to establish diagnostic reliability. Inter-rater agreement in PTSD diagnosis, reflected by the kappa statistic, was acceptable for PTSD (k = .81). As reviewed by Weathers, Keane, and Davidson (2001), the CAPS has excellent support for its reliability, with alpha coefficients generally ranging from .64 to .88. As noted in Table 1, alpha coefficients in the current sample ranged from .64 to .76, which is consistent with previous investigations of the interview’s internal consistency involving clinical samples (Weathers et al., 2001). Two to three day test–retest reliability was found to range from .78 to .87 (Weathers et al., 2001). Of particular importance in this study, the CAPS has been shown to be sensitive to the detection of PTSD in individuals following a MVA (Blanchard & Hickling, 1997).

Participants completed two additional self-report scales, the original Impact of Event Scale (IES, Horowitz et al., 1979) and the PTSD Symptom Scale-Self Report (PSS-SR; Foa, Riggs, Dancu, & Rothbaum, 1993). The IES contains 15 items that are rated on a 0–3 Likert Scale ranging from “not at all” to “often” and are distributed across two subscales that assess intrusion (7 items, e.g., “I thought about it when I didn’t mean to”) and avoidance (8 items, e.g., “I stayed away from reminders of it”). The IES has been shown to have high internal consistency with alpha coefficients of .78 for the intrusion subscale and .82 for the avoidance subscale in a sample of outpatients (Horowitz et al., 1979). Split-half reliability of the total scale was .86 and the 1-week test–retest reliability was .89 for the intrusion subscale and .79 for the avoidance subscale (Horowitz et al., 1979). The PSS-SR contains 17 items, reflecting the DSM-IV symptoms of PTSD, which are rated on a 0–3 Likert Scale, ranging from “not at all” to “5 or more times/week-almost always”. Items are summed to yield a total score. Foa et al. (1993) evaluated the psychometric properties of the PSS-SR with 46 female rape victims and 72 female non-sexual assault victims. In this sample, the PSS-SR showed high internal consistency (α = .91) and good 1-month test–retest reliability (r = .74). Convergent validity of the PSS-SR, with the IES and State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970) also was demonstrated, with correlations ranging from .52 to .81 (Foa et al., 1993). Higher scores on both of these measures indicate the presence of more PTSD symptoms. Alpha coefficients of the CAPS, IES, and PSS-SR for the current sample are reported in Table 1.
Table 1
Means, standard deviations, alpha coefficients, and zero-order correlations among Impact of Event Scale-Revised (IES-R) subscales and measures of Posttraumatic Stress Disorder, Anxiety, and Depression (partial correlations controlling for litigation status are shown in parentheses)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Intrusion</th>
<th>Avoidance</th>
<th>Hyperarousal</th>
<th>Mean</th>
<th>S.D.</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>IES-R</td>
<td></td>
<td></td>
<td></td>
<td>1.57</td>
<td>.99</td>
<td>.90</td>
</tr>
<tr>
<td>Intrusion</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoidance</td>
<td>.74</td>
<td>–</td>
<td></td>
<td>1.44</td>
<td>.90</td>
<td>.86</td>
</tr>
<tr>
<td>Hyperarousal</td>
<td>.86</td>
<td>.71</td>
<td>–</td>
<td>1.81</td>
<td>1.07</td>
<td>.85</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>1.59</td>
<td>.90</td>
<td>.95</td>
</tr>
</tbody>
</table>

PTSD measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Intrusion</th>
<th>Avoidance</th>
<th>Hyperarousal</th>
<th>Mean</th>
<th>S.D.</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>IES-Intrusion</td>
<td>.86 (.84)</td>
<td>.66 (.64)</td>
<td>.75 (.70)</td>
<td>16.91</td>
<td>.65</td>
<td>.91</td>
</tr>
<tr>
<td>IES-Avoidance</td>
<td>.63 (.61)</td>
<td>.76 (.75)</td>
<td>.61 (.59)</td>
<td>16.35</td>
<td>.93</td>
<td>.87</td>
</tr>
<tr>
<td>IES-Total</td>
<td>.81 (.80)</td>
<td>.77 (.76)</td>
<td>.74 (.73)</td>
<td>33.26</td>
<td>18.98</td>
<td>.93</td>
</tr>
<tr>
<td>CAPS-Re-experiencing</td>
<td>.66 (.62)</td>
<td>.47 (.44)</td>
<td>.57 (.53)</td>
<td>14.54</td>
<td>8.58</td>
<td>.76</td>
</tr>
<tr>
<td>CAPS-Avoidance and Numbing</td>
<td>.53 (.50)</td>
<td>.48 (.46)</td>
<td>.49 (.48)</td>
<td>16.02</td>
<td>9.95</td>
<td>.67</td>
</tr>
<tr>
<td>CAPS-Hyperarousal</td>
<td>.50 (.49)</td>
<td>.39 (.37)</td>
<td>.52 (.49)</td>
<td>13.93</td>
<td>7.74</td>
<td>.64</td>
</tr>
<tr>
<td>PSS-SR-Re-Experiencing</td>
<td>.86 (.83)</td>
<td>.63 (.62)</td>
<td>.75 (.71)</td>
<td>5.75</td>
<td>4.11</td>
<td>.87</td>
</tr>
<tr>
<td>PSS-SR-Avoidance and Numbing</td>
<td>.69 (.66)</td>
<td>.73 (.71)</td>
<td>.71 (.69)</td>
<td>9.64</td>
<td>5.42</td>
<td>.81</td>
</tr>
<tr>
<td>PSS-SR-Hyperarousal</td>
<td>.75 (.71)</td>
<td>.60 (.57)</td>
<td>.85 (.83)</td>
<td>7.75</td>
<td>4.14</td>
<td>.79</td>
</tr>
</tbody>
</table>

Anxiety measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>STAI-State</th>
<th>STAI-Trait</th>
<th>BAI</th>
<th>Mean</th>
<th>S.D.</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAI-State</td>
<td>.58 (.58)</td>
<td>.45 (.43)</td>
<td>.61 (.63)</td>
<td>46.90</td>
<td>15.64</td>
<td>.96</td>
</tr>
<tr>
<td>STAI-Trait</td>
<td>.58 (.59)</td>
<td>.52 (.49)</td>
<td>.64 (.66)</td>
<td>47.70</td>
<td>13.28</td>
<td>.95</td>
</tr>
<tr>
<td>BAI</td>
<td>.63 (.61)</td>
<td>.58 (.56)</td>
<td>.70 (.70)</td>
<td>18.70</td>
<td>12.52</td>
<td>.93</td>
</tr>
</tbody>
</table>

Depression measure

<table>
<thead>
<tr>
<th>Measure</th>
<th>BDI-II</th>
<th>M–C</th>
<th>Mean</th>
<th>S.D.</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDI-II</td>
<td>.65 (.63)</td>
<td>.53 (.51)</td>
<td>.72 (.71)</td>
<td>20.65</td>
<td>11.95</td>
</tr>
</tbody>
</table>

Social desirability measure

<table>
<thead>
<tr>
<th>Measure</th>
<th>M–C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.05</td>
</tr>
</tbody>
</table>

Note: Values listed in bold are significant \((p < .05)\). CAPS: Clinician-Administered PTSD Scale; IES: Impact of Event Scale; PSS-SR: PTSD Symptom Scale-Self Report; STAI: State-Trait Anxiety Inventory; BAI: Beck Anxiety Inventory; BDI: Beck Depression Inventory; M–C: Marlowe–Crowne Social Desirability Scale.

2.2.3. Anxiety measures

The State-Trait Anxiety Inventory (STAI, Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) is a well-recognized questionnaire, comprised of a 20-item state (STAI-S) and a 20-item trait anxiety (STAI-T) subscale. The STAI was administered to examine the convergent validity of the IES-R. Good internal consistency has been reported for the STAI \((\alpha \text{ ranging from .86 to .95; Spielberg et al., 1983})\) and the scale’s test–retest reliability ranges from .75 to .86 for intervals 30 days or less (Spielberger et al., 1983). Convergent validity of the STAI-T and other measures of trait anxiety has been noted (e.g., Bieling, Antony, & Swinson, 1998). Higher scores on both subscales indicate the presence of more anxiety.

The Beck Anxiety Inventory (BAI, Beck, Epstein, Brown, & Steer, 1988) also was used to assess anxiety. The BAI is a 21-item self-report inventory of common symptoms of anxiety (e.g., difficulty breathing, feelings of choking) that was developed to assess anxiety severity while minimizing overlap with assessment of depression. Participants rate the extent to which they were bothered by each item during the past week on a 0–3 scale ranging from “not at all” to “severely.” The inventory has been shown to have high internal consistency \((\alpha = .92)\) and test–retest reliability \((r = .75)\) in a sample of psychiatric outpatients (Beck et al., 1988). Convergent validity of the BAI and other measures of anxiety, such as the Hamilton Rating Scale for Anxiety and diary ratings of anxiety, has been noted (Beck et al., 1988; deBeurs, Wilson, Chambless, Goldstein, & Feske, 1997). Alpha coefficients for the STAI and the BAI in the current sample are reported in Table 1.

2.2.4. Depression measure

The Beck Depression Inventory-II (BDI-II, Beck, Steer, & Brown, 1996) was administered to examine the convergent validity of the IES-R. The 21-items of this scale are rated on a 0–3 scale to evaluate current depressive symptoms, with higher scores indicating greater depression. One-week test–retest reliability was demonstrated to be .93 among outpatients (Beck et al., 1996) and the scale appears to be highly reliable with an alpha coefficient of .93 for college samples and .92 for psychiatric samples. The alpha coefficient of the BDI-II
for the current sample is reported in Table 1. The BDI-II has been shown to correlate highly with other measures of depression (Steer & Clark, 1997), supporting its construct validity. In considering the use of the BDI-II with this sample, where chronic pain is a common co-morbid condition (e.g., Blanchard et al., 1995), one choice is to remove somatic items when scoring the instrument. However, this procedure did not improve accuracy of an earlier version of the scale (e.g., Geisser, Roth, & Robinson, 1997). To facilitate comparison of the current data with previously published studies, the original scoring method was used.

2.2.5. Social desirability measure

The Marlowe–Crowne Social Desirability Scale (M–C) was developed to assess the tendency to self-report in a socially desirable manner (Crowne & Marlowe, 1960). The M–C is a 33-item questionnaire, answered in a true/false format. Based on the rationale for the Lie Scale of the MMPI (Meehl & Hathaway, 1946), the items for this measure were chosen from a set of behaviors that are culturally sanctioned but for which total compliance is improbable. Higher scores on the measure indicate increased defensiveness. In a sample of 10 male and 29 female undergraduates, the M–C was found to have good 1-month test–retest reliability ($r = .89$) and an internal consistency coefficient of .88 (Crowne & Marlowe, 1960). The alpha coefficient of the M–C for the current sample is reported in Table 1.

2.3. Procedure

Procedures were reviewed by the Institutional Review Board at The University at Buffalo-SUNY. Participants were interviewed individually and then completed the questionnaires.

2.4. Analytic strategy

In the first step, the IES-R was submitted to confirmatory factor analysis (CFA), to determine if the three-factor structure (Intrusion, Avoidance, Hyperarousal) reported by Weiss and Marmar (1997) would be replicated. Maximum Likelihood Estimation was employed, using MPlus software (Muthén & Muthén, 2004). Although traditional guidelines would suggest that the sample size might be too small for a CFA, recent discussion by MacCallum, Widaman, Zhang, and Hong (1999) suggests otherwise. These authors argue against the use of invariant rules for determination of sample size for a CFA and instead, present a mathematical framework for computing the necessary sample size based on the size and variability of the obtained communalities and whether factors are overdetermined. In the case of the present data, with five or more items per factor, moderate to high expected communalities based on past exploratory factor analyses, and $n = 182$, the approximate value of the congruence coefficient (which reflects the degree of correspondence between sample and population factors) is .98, which is excellent within MacCallum et al. (1999) framework. Thus, the sample size in the present report is adequate for a CFA. Coefficient alpha was computed for each obtained subscale.

The second data analytic step involved provision of descriptive statistics and examination of intercorrelations among the obtained subscales to determine overlap in variance. Next, the IES-R was compared with the other PTSD measures, including the IES, to examine convergent validity. Consideration of zero-order correlations with related measures of anxiety and depression also was included in this step. These analyses were repeated, controlling for the presence of litigation, to examine if this factor influenced these associations. Next, correlations between the obtained subscales and the M–C were computed, to examine associations with social desirability. The last analytic step involved comparison of individuals with and without diagnosable PTSD with respect to the IES-R subscales, in an effort to examine discriminative validity. As part of this analytic step, a discriminant function analysis was included to examine the extent to which the IES-R could categorize correctly participants with and without PTSD and to determine which subscale(s) contributed to this classification.

3. Results

3.1. How well does the factor structure of the IES-R approximate the tripartite model of DSM?

A CFA was conducted to investigate whether the IES-R fit well with the three-factor model (Weiss & Marmar, 1997; Intrusion, Avoidance, and Hyperarousal). To ensure that the model was identified, one parameter estimate of each latent variable was constrained to equal 1 (Kline, 1998). Factor correlations were freely estimated and there were no cross-loadings. Several fit indices were used in examining the fit of the model, following the guidelines suggested by Marsh, Hau, and Wen (2004). These guidelines indicate that adequate fit is demonstrated if the comparative fit index (CFI) and Tucker Lewis Index (TLI) are close to .90, the root-mean square error of approximation (RMSEA) is
close to .08, and the standardized root mean squared residual (SRMR) is close to .06. As is appropriate for factor models with simple structures (in which many parameter estimates will be constrained to zero), we applied guidelines that are slightly less stringent than those suggested by Hu and Bentler (1999). Kline (1998) suggests that the $\chi^2$/df also should be used in evaluating fit of CFAs, in which a $\chi^2$/df of less than 3.0 suggests that the model is fitting the data well. Additionally, the size of the factor loadings was considered when determining model fit (Byrne, 2001; Kline, 1998), using criteria described by Comrey and Lee (1992) which suggests that loadings of .32 are substantial (but poor), and loadings greater than .45 are substantial (and fair).

The results of the initial CFA suggested that the proposed factor structure did not fit the data well, $\chi^2(206, N = 182) = 591.517, p < .001$; $\chi^2$/df = 2.87; CFI = .84; TLI = .82; RMSEA = .10, SRMR = .07. Modification indices indicated that freeing error covariances might improve the fit of the model; those covariances that made theoretical sense (e.g., among constructs that were conceptually related to one another) were added to the model (items 15 and 2; items 11 and 17, see Table 2) and the fit of the model was reevaluated. The revised model provided a more adequate fit to the data, $\chi^2(204, N = 182) = 476.477, p < .001$; $\chi^2$/df = 2.34; CFI = .89; TLI = .87; RMSEA = .09, SRMR = .06. Examination of the modification indices suggested that none exhibited cross-loadings. Furthermore, all of the factor loadings loaded substantially onto their respective factors ($> .45$). Therefore, based on the revised model’s fit indices, the size of the factor loadings, and substantive concerns (e.g., the structure of PTSD as defined within the DSM), we chose to keep this structure as the final model. Table 2 displays the standardized factor loadings, communalities, and critical ratios for the final model.

3.2. Internal consistency

Alpha coefficients for each subscale are shown in Table 1 and appear adequate.

3.3. Concurrent validity

3.3.1. Descriptive statistics and within scale correlations

Means and standard deviations for each of the measures are presented in Table 1. Pearson correlations were computed to examine associations within the obtained subscales of the IES-R. As noted in Table 1, the observed correlation between each pair of subscales was high (range .71–.86). These associations are not significantly different (all $p$’s > .05) from those reported by Creamer et al. (2003).

3.3.2. Correlation with measures of PTSD, anxiety, and depression

For the PTSD measures, it was hypothesized that cross-measure correlations for a specific subscale (e.g., hyperarousal) would be significantly higher than cross-measure correlations between two different subscales (e.g., hyperarousal and intrusion). This hypothesis was pursued using the procedure outlined by Meng, Rosenthal, and Rubin (1992) to compare correlated correlation coefficients.

In comparing the IES-R with the original IES (see Table 1), the correlation between the two Intrusion subscales ($r = .86$) was significantly higher than the correlation between the IES-R Avoidance subscale and the IES Intrusion subscale ($r = .66, Z = 6.59, p < .000001$) and the correlation between the IES-R Hyperarousal subscale and the IES Intrusion subscale ($r = .75, Z = 1.77, p < .05$), suggesting some degree of concurrent validity in the subscales. Likewise, the correlation between the original and revised Avoidance subscales ($r = .76$) was significantly higher than the correlation between the IES Avoidance subscale and the IES-R Intrusion subscale ($r = .63, Z = 1.84, p < .05$) and the correlation between the IES Avoidance subscale and the IES-R Hyperarousal subscale ($r = .61, Z = 2.07, p < .01$). Because some authors rely on the IES-Total score, correlations between the IES-R subscales and this score are shown in Table 1 as well.

Similar analyses were conducted with the IES-R and the CAPS. The correlation between the CAPS Re-experiencing subscale and the IES-R Intrusion subscale ($r = .66$) was significantly higher than the correlations between both the CAPS Re-experiencing and IES-R Avoidance subscales ($r = .47, Z = 4.43, p < .0001$) and IES-R Hyperarousal subscales ($r = .57, Z = 2.96, p < .001$). However, a similar pattern of cross-measure correlation was not noted with the IES-R Avoidance subscales. In this instance, the correlations between the CAPS Avoidance and numbing subscale and the IES-R Avoidance subscale ($r = .48$) was not significantly different from correlations between the CAPS Avoidance and numbing subscale and the other two IES-R subscales (IES-R Intrusion $r = .53$, IES-R Hyperarousal $r = .49$). Although the correlation between the IES-R

---

1 A copy of the complete correlation matrix is available from the first author.
and CAPS Hyperarousal subscales \((r = .52)\) was significantly higher than that noted between the IES-R Avoidance subscale and the CAPS Hyperarousal subscale \((r = .39, Z = 11.93)\), a non-significant difference was noted when comparing the correlation between the IES-R Hyperarousal/CAPS Hyperarousal subscales with the IES-R Intrusion/CAPS Hyperarousal subscales \((r = .50)\).

Analyses between the correlations of the IES-R and the PSS-SR subscales showed similar results. The correlation between the IES-R Intrusion and PSS-SR Re-experiencing subscales \((r = .86)\) was significantly higher than the correlations between the IES-R Avoidance and PSS-SR Re-experiencing subscales \((r = .63, Z = -7.36, p < .00001)\) and PSS-SR Hyperarousal subscales \((r = .75, Z = 4.59, p < .0003)\). Across measures, the results suggest that the IES-R intrusion subscale, and to a lesser extent, the IES-R Hyperarousal subscale, show some degree of specificity across symptom clusters.

For the anxiety and depression measures, it was hypothesized that each of the IES-R subscales would show significant correlations with these constructs. As noted in Table 1, this hypothesis was supported (all \(p\)'s < .05).

### 3.3.3. Correlations controlling for litigation status

In order to examine whether scores on the IES-R were influenced by the presence of MVA-related litigation, partial correlations between each subscale and measures of PTSD, anxiety, and depression were computed, controlling for litigation status. The pattern and magnitude of the partial correlations was similar to the bivariate associations (see Table 1).
3.4. Social desirability

The correlation between each subscale of the IES-R and the M–C was non-significant (see Table 1), suggesting that social desirability was not highly associated with the subscales.

3.5. Discriminative validity: differences between individuals with and without PTSD

To examine whether the subscales of the IES-R had discriminative validity in distinguishing between individuals with and without PTSD, a collection of analysis of variance and covariance were conducted. These analyses compared individuals with PTSD (PTSD+, \(n = 98\)) to those without (PTSD−, \(n = 84\)) on each of the IES-R subscales. As noted in Table 3, the two groups reported significantly different scores on each subscale (Intrusion, \(t(180) = -5.88, p < .0001, 16.1\%\) variance accounted for; Avoidance, \(t(180) = -4.18, p < .0001, 8.9\%\) variance; Hyperarousal, \(t(180) = -6.16, p < .0001, 17.4\%\) variance). These analyses were repeated using the presence of pain complaints as a covariate, as pain and PTSD symptoms appear to interact synergistically in their contributions to psychological distress (e.g., Palyo & Beck, 2005; Sharp & Harvey, 2001) and the two groups differed significantly in the proportion of individuals with pain complaints. Controlling for the presence of pain complaints, the results continued to show significant between-group differences on each subscale of the IES-R (Intrusion: \(F(1,181) = 24.6, p < .0001, 12\%\) variance; Avoidance: \(F(1,181) = 11.6, p < .01, 6\%\) variance; Hyperarousal: \(F(1,181) = 26.8, p < .0001, 13\%\) variance) with the PTSD+ group scoring significantly higher than the PTSD− group. Owing to a significant difference between the two groups with respect to time since MVA (\(t = 1.99, p < .05\)), a second ANCOVA was conducted, using months since MVA as a covariate, with the PTSD+ group again scoring significantly higher than the PTSD− group on each subscale of the IES-R (Intrusion: \(F(1,181) = 30.7, p < .0001, 15\%\) variance; Avoidance: \(F(1,181) = 15.4, p < .0001, 8\%\) variance; Hyperarousal: \(F(1,181) = 33.8, p < .0001, 16\%\) variance).

Lastly, a discriminant function analysis (DFA) was conducted to examine the specificity and sensitivity of the IES-R subscales in identifying individuals with and without PTSD. The three subscales loaded on one function which correctly classified 69.2% of the sample (Wilks’ lambda = .82, \(\chi^2(3, N = 181) = 35.92, p < .0001\)). Sensitivity was 74.5 and specificity was 63.1. The structure matrix of correlations between discriminating variables and the one discriminant function suggested that the Hyperarousal (\(r = .97\)) and the Intrusion (\(r = .93\)) subscales provided the best contributions to the obtained function, relative to the Avoidance subscale (\(r = .66\)). By way of comparison, a similar DFA was conducted with the original IES subscales. The two subscales loaded on one function, which correctly classified 70.3% of the sample (Wilks’ lambda = .81, \(\chi^2(2, N = 181) = 37.52, p < .0001\)). Sensitivity was 77.6 and specificity was 61.9. The structure matrix of correlations suggested that the Intrusion (\(r = .97\)) subscale provided a better contribution to the obtained function, relative to the Avoidance subscale (\(r = .69\)). Thus, the original and revised IES-R appeared similar with respect to sensitivity and specificity.

4. Discussion

This study examined the factor structure, internal consistency, concurrent validity, the influence of social desirability, and discriminative validity of the IES-R in a sample of individuals who had experienced a serious MVA. Results supported the three-factor structure of the

| Table 3 Demographic information and Impact of Event Scale-Revised (IES-R) scores for participants with diagnosable post-traumatic stress disorder (PTSD+) and no PTSD (PTSD−) |
|-----------------|-----------------|-----------------|
| Age | 42.9 (10.98) | 42.7 (13.84) |
| Race (%Caucasian) | 81.6% | 84.5% |
| Education (%Associates degree or higher) | 54.8% | 43.9% |
| Gender (%female) | 82.6% | 72.6% |
| Presence of pain complaints (%present) | 83% | 62% |
| Months since MVA | 28.3 (65.14) | 53.6 (99.29) |
| IES-R Intrusion | 1.9 (.85) | 1.1 (.88) |
| IES-R Avoidance | 1.7 (.85) | 1.2 (.88) |
| IES-R Hyperarousal | 2.2 (.94) | 1.3 (1.02) |
IES-R, Intrusion, Avoidance, and Hyperarousal, with adequate internal consistency noted for each subscale. As with previous accounts (e.g., Creamer et al., 2003), the three subscales of the IES-R showed a high degree of intercorrelation. Correlations between corresponding subscales on the IES-R and the original IES were high and specific. In particular, cross-measure correlations for a specific subscale were significantly higher than cross-measure correlations between two different subscales, providing support for the concurrent validity of both forms of this measure. Examination of correlations between the IES-R and other measures of PTSD were high but somewhat less specific. In comparison with the CAPS and the PSS-SR, the IES-R Intrusion subscale, and to a lesser extent, the Hyperarousal subscale, showed some degree of specific cross-measure correlation across symptom clusters. The IES-R Avoidance subscale failed to show specificity. These relationships appeared similar across measurement modality, although correlations between two self-report scales (IES-R and PSS-SR) were understandably higher than cross-modality correlations (IES-R and CAPS). Each of the three subscales showed significant associations with widely used measures of anxiety and depression, as predicted. The obtained correlations between the IES-R subscales and measures of PTSD, anxiety, and depression did not change notably when litigation status was controlled. None of the subscales showed significant correlations with social desirability. Higher scores for PTSD+ participants on each IES-R subscale, relative to non-PTSD participants, offers support for the discriminative validity, particularly as these differences were maintained even when controlling for the influence of pain complaints and elapsed time since the MVA. A discriminant function analysis indicated that the obtained subscales in combination were able to correctly classify 69% of the sample. Sensitivity was acceptable, although specificity was somewhat lower. Closer examination indicated that only the Intrusion and Hyperarousal subscales appeared salient in this equation.

Taken together, these results extend psychometric support for the IES-R. The three-factor structure was supported in this investigation, unlike the findings of Creamer et al. (2003), suggesting that the IES-R is not simply a measure of general distress. It is possible that differences in methodology account for the discrepancy between these studies. For example, Creamer used markedly different recruitment strategies to enroll clinical and non-clinical participants, possibly resulting in heterogeneity within the total sample in this study. This heterogeneity conceivably could explain the single factor solution noted by these authors, which might reflect general distress. The current report, in contrast, relied upon a more homogeneous sample, ensuring greater clarity in interpretation of the obtained factor structure. Another feature that strengthens the current report is the use of stringent diagnostic procedures in defining PTSD+ and PTSD– subsamples, which aids interpretation of the discriminative validity data.

Importantly, the IES-R appears to have some specific cross-measure agreement in the assessment of PTSD symptomatology. The IES-R Intrusion subscale showed significantly higher within-scale correlations with matching subscales on the CAPS and the PSS-SR, relative to between-scale correlations with these measures. The IES-R Hyperarousal subscale showed some degree of cross-measure agreement, as well. Surprisingly, the IES-R Avoidance subscale did not perform well in its subscale associations with the CAPS and PSS-SR. In particular, this subscale did not show higher within-scale correlations, relative to its between-scale correlations, suggesting a lack of specific cross-measure agreement in its content. Examination of this subscale in Table 2 indicates that many of the items reflect active avoidance, with less item content reflecting emotional numbing. In fact, several numbing symptoms (diminished interest, detachment from others, and a sense of foreshortened future) are not represented in the item content of the IES-R. In contrast, both the CAPS and the PSS-SR map closely onto the DSM criteria, wherein avoidance and numbing symptoms are grouped together in the same cluster. As such, the lower level of specific cross-level agreement could reflect differences in the amount of item content that describes emotional numbing in the IES-R. Alternatively, the IES-R Avoidance subscale may need additional development if it is to converge more closely with the DSM. Interestingly, this subscale contributed the least to the discriminant function that was developed to differentiate individuals with and without PTSD. These findings suggest that the IES-R Avoidance subscale does not fit as closely with the DSM criteria for PTSD, relative to the other two subscales. Given on-going discussion in the literature concerning the underlying factor structure of PTSD (e.g., Asmundson et al., 2000; DuHamel et al., 2004), it is possible that the IES-R can augment diagnostically oriented measures, such as the CAPS and PSS-SR, in the assessment of PTSD. In particular, the IES-R Avoidance subscale may provide a more focused assessment of effortful avoidance, relative to measures that combine avoidance and numbing.

Although the IES-R was not developed as a diagnostic tool, examination of its discriminative
validity suggests that the measure can differentiate between individuals with and without PTSD. However, the specificity and sensitivity of the obtained discriminant function indicates that the scale does not perform optimally as a screening tool. To examine this issue, Creamer and colleagues (2003) compared the IES-R with the PTSD Checklist (PCL, Weathers, Litz, Herman, Huska, & Keane, 1993). Using a PCL cut-off score of 50 to identify PTSD+ cases, Creamer et al. report that a total score of 33 on the IES-R yielded diagnostic sensitivity of .91 and specificity of .82. On reflection, one may wonder if there are any advantages to this use of the IES-R. Because the measure does not correspond exactly to DSM-based definitions of PTSD and currently contains more items than typical screening measures, this may not be the optimal use for the IES-R. Future studies with this scale might wish to consider this issue.

As with most empirical studies, this work has certain limitations. First, the results may not generalize beyond individuals who have experienced a serious MVA and it is possible that our findings are unique to survivors of this type of trauma. Given our objective of providing psychometric validation of the IES-R, a homogeneous sample represents a good methodological choice. An important next step will involve replication with survivors of a more diverse collection of traumas. Second, modification indices were used to free error covariances in the CFA. Although theoretically justified in light of considerable overlap between items, this technique should be replicated with an independent sample. A third limitation of this study lies in its demographic homogeneity. The majority of the sample was female and Caucasian, precluding invariance testing of confirmatory models across gender and ethnicity. In light of recent data which suggest that cultural influences interact with gender differences in PTSD symptomatology (e.g., Norris, Perilla, Ibanez, & Murphy, 2001), future work should examine whether the IES-R factor structure is invariant across culture and gender. This research direction is facilitated by the presence of translated versions of the scale (e.g., Asukai et al., 2002; Baguena et al., 2001; Wu & Chan, 2003). Lastly, it will be useful for future studies to look at other parameters of the IES-R, including the scale’s sensitivity to changes produced by treatment.

In sum, the report extends the available psychometric support for the IES-R. In light of the established history of the IES, this study suggests that the revised version of this scale may enjoy a similar trajectory. Although the scale mirrors the basic structure of PTSD as defined within current diagnostic systems, it retains an emphasis on assessing the basic phenomena of intrusion and avoidance that characterized Horowitz’s original purpose for the scale. It appears to be a sound measure for both clinical and research purposes. Future studies would be advised to examine cross-cultural and gender effects, in order to further explore the measurement parameters of this scale.

Acknowledgements

We appreciate the assistance and cooperation of our participants, without whom this work would not be possible.

This work was supported in part by a grant from the National Institute of Mental Health (MH64777).

References


