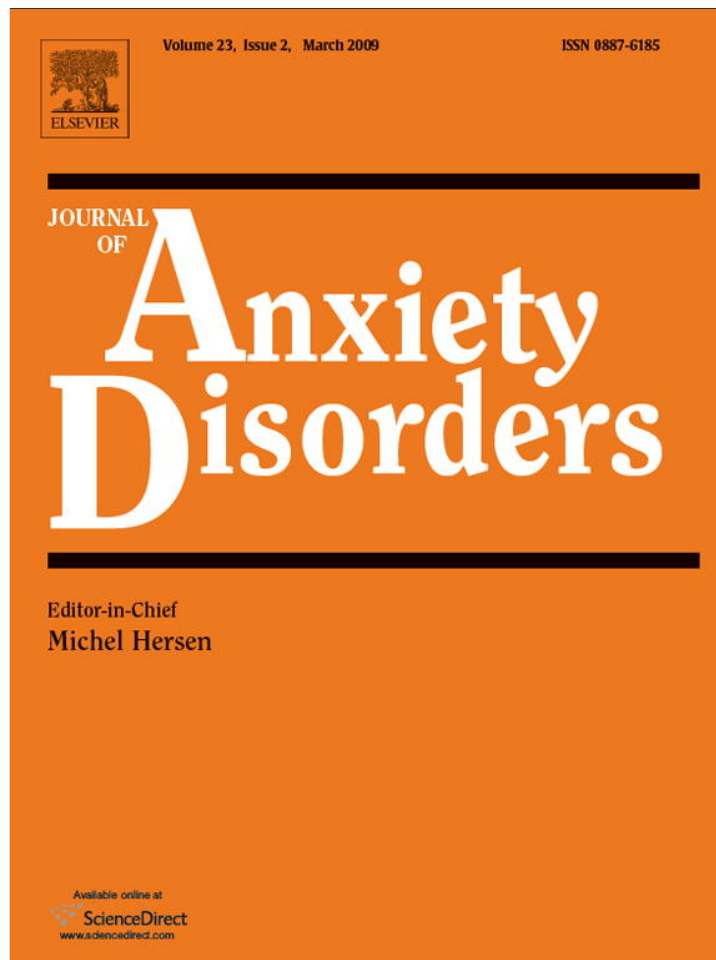


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

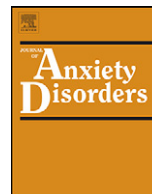
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Anxiety Disorders



## The validity and diagnostic efficiency of the Davidson Trauma Scale in military veterans who have served since September 11th, 2001

Scott D. McDonald<sup>a,\*</sup>, Jean C. Beckham<sup>a,b,c,d</sup>, Rajendra A. Morey<sup>a,c,d</sup>, Patrick S. Calhoun<sup>a,b,c,d</sup>

<sup>a</sup> VA Mid-Atlantic Region Mental Illness Research, Education and Clinical Center (MIRECC), Durham, North Carolina, United States

<sup>b</sup> VA Center for Health Services Research in Primary Care, Durham VA Medical Center, Durham, North Carolina, United States

<sup>c</sup> Durham VA Medical Center, Durham, North Carolina, United States

<sup>d</sup> Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina, United States

### ARTICLE INFO

#### Article history:

Received 29 February 2008

Received in revised form 14 July 2008

Accepted 25 July 2008

#### Keywords:

Posttraumatic stress disorder

Military veterans

Test validity

Differential diagnosis

Diagnostic efficiency

### ABSTRACT

The present study examined the psychometric properties and diagnostic efficiency of the Davidson Trauma Scale (DTS), a self-report measure of posttraumatic stress disorder (PTSD) symptoms. Participants included 158 U.S. military veterans who have served since September 11, 2001 (post-9/11). Results support the DTS as a valid self-report measure of PTSD symptoms. The DTS demonstrated good internal consistency, concurrent validity, and convergent and divergent validity. Diagnostic efficiency was excellent when discriminating between veterans with PTSD and veterans with no Axis I diagnosis. However, although satisfactory by conventional standards, efficiency was substantially attenuated when discriminating between PTSD and other Axis I diagnoses. Thus, results illustrate that potency of the DTS as a diagnostic aid was highly dependent on the comparison group used for analyses. Results are discussed in terms of applications to clinical practice and research.

Published by Elsevier Ltd.

### 1. Introduction

Self-report measures are valuable tools for clinicians and researchers, as they are quick and cost-effective methods for assessing symptoms associated with mental illness. In the last two decades, several self-report measures of posttraumatic stress disorder (PTSD) have been developed (see Brewin, 2005; Norris & Hamblen, 2004; for reviews). Concurrently, there has been a growing appreciation for the reality that for a measure to have utility, it is essential that support for its validity has been demonstrated (Hunsley & Mash, 2005). As evidence accumulates for the negative impact of PTSD on overall health (Beckham et al., 1998; Dohrenwend et al., 2007; Taft, Stern, King, & King, 1999), family adjustment (Jordan et al., 1992) and health care costs (Walker et al., 2003) the need for brief and valid measures of PTSD symptoms has become clear.

Prior studies have validated various PTSD symptom questionnaires for use with several targeted groups, including breast cancer patients (Andrykowski, Cordova, Studts, & Miller, 1998),

crime victims (Wohlfarth, van den Brink, Winkel, & ter Smitten, 2003), Vietnam-era combat veterans (Forbes, Creamer, & Biddle, 2001), female veterans in primary care (Dobie et al., 2002; Lang, Laffaye, Satz, Dresselhaus, & Stein, 2003) and older adults in primary care (Cook, Elhai, & Areán, 2005). Replications such as these are important, as a symptom scale may take on different properties in different populations (Bossuyt et al., 2003; Brewin, 2005). For example, Blanchard et al. (1996) found that a score of 44 or higher on the PCL Checklist was most effective at identifying PTSD while minimizing false positives in a sample of motor vehicle accident and sexual assault victims. In contrast, Lang et al. (2003), using the same measure, found that a score in the range of 28–30 was most effective in detecting PTSD in female veterans who visited a primary care clinic.

Unfortunately, many populations that are at high-risk for trauma exposure do not have adequately-validated measures available. For example, in spite of the increasing need for valid PTSD screening instruments for returning military service personnel (Hoge et al., 2004), no self-report measure of PTSD has yet been validated with veterans who have served since September 11th, 2001 (post-9/11). It is estimated that 35% of OIF veterans will access mental health services in the year after returning home, and 5–20% will meet criteria for PTSD (Hoge, Auchterlonie, & Milliken, 2006; Hoge et al., 2004). As there are now several empirically-supported treatments for PTSD (American Psychiatric Association,

\* Corresponding author. Present address: Defense and Veterans Brain Injury Center, Richmond Veterans Affairs Medical Center, Research (151), 1201 Broad Rock Boulevard, Richmond, VA 23249 United States. Tel.: +1 804 675 5000x3633.

E-mail address: Scott.McDonald@va.gov (S.D. McDonald).

2004; Bisson & Andrew, 2005; Department of Veterans Affairs & Department of Defense, 2004), there is compelling incentive to validate screening tools for the identification of PTSD in high-risk groups, such as military personnel and veterans.

Diagnostic tests are most efficient when a group identified as having the condition is compared to an equal number of those that exhibit none of the clinical characteristics of the condition, e.g. healthy controls. It is important to consider that clinical characteristics of the comparison group may have significant effects on the efficiency of the test in question, and thus their generalizability (Coyne & Thompson, 2007; Streiner, 2003). For example, individuals with Major Depressive Disorder endorse many of the symptoms that are found in PTSD (e.g., poor concentration, sleep difficulties, and anhedonia), and thus tend to score higher on PTSD symptom questionnaires than healthy controls (Shalev et al., 1998). In effect, a score that is very efficient when discriminating between PTSD and healthy controls may be less efficient in discriminating between PTSD and individuals with other presenting problems. The latter scenario more approximates conditions in a mental health clinic, in which most patients will be in distress and the clinician is faced with the often challenging task of differential diagnosis (cf. Hankin, Spiro, Miller, & Kazis, 1999). Unfortunately, prior studies have rarely described or assessed the clinical characteristics of their comparison groups. Thus, the literature provides little evidence for the diagnostic efficiency of PTSD symptom questionnaires in a mental health setting.

The purpose of the current study was to examine the validity and diagnostic efficiency of the Davidson Trauma Scale (DTS; Davidson, Book, et al., 1997) in a group of veterans who served after September 11th, 2001. The DTS is a self-report measure of the 17 PTSD symptoms as described in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR; American Psychiatric Association, 2000). As a diagnostic tool, Davidson, Book, et al. (1997) demonstrated that the DTS performed well at discriminating 67 individuals with PTSD from 62 without PTSD (area under the curve [AUC] = 0.88, S.E. = 0.02) using a semi-structured interview (SCID; Spitzer, Williams, Gibbon, & First, 1990) as the reference standard. A DTS score of 40 was recommended as the optimal cut-point for accurate classification of those with or without PTSD (efficiency = 0.83). This cut-point correctly classified 69% of individuals with PTSD (sensitivity = 0.69) and 95% of those who did not have PTSD (specificity = 0.95).

No previous studies have examined the psychometric properties of the DTS in veterans who have served post-9/11. Like most PTSD screening measures, the ability of the DTS to discriminate between those with PTSD and other psychiatric disorders is unknown. Therefore, the current study tested the ability of the DTS to discriminate between veterans with PTSD and two comparison groups: (1) veterans with no Axis I diagnosis and (2) veterans without PTSD but with a current diagnosis of another Axis I disorder.

## 2. Method

### 2.1. Participants and procedures

The sample consisted of 226 volunteer participants in the Mid-Atlantic Mental Illness Research, Education and Clinical Center (MIRECC) Recruitment Database for the Study of Post-Deployment Mental Health. Participants were veterans who have served in the United States Armed Forces since September 11, 2001. About half (53%) of the participants had been stationed in a region of conflict in support of Operation Enduring Freedom or Operation Iraqi Freedom. Participants were recruited from four VISN-6 Veterans Affairs medical centers through mailings advertisements, and

clinician referrals. Informed consent was obtained after explaining procedures. The veterans were administered a battery of questionnaires related to post-deployment mental illness, including: psychiatric symptoms, mental-health service utilization, health, and health-related behaviors. Diagnosis was established by the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I/P; First, Spitzer, Gibbon, & Williams, 1994), a semi-structured interview administered by trained masters- or Ph.D.-level clinicians. Participants in this sample were among those used for an evaluation of the factor structure of the DTS, which is presented elsewhere (McDonald et al., 2008).

Of the initial pool of 226 participants, 158 veterans recorded a traumatic event on the DTS that clearly met DSM-IV Criterion A1 (i.e., “involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others”; APA, 2000, p. 467). Of those 158 veterans, 75 (47%) recorded war zone-related traumas (e.g., “my truck hit by an improvised explosive device”) and 83 (53%) recorded traumas occurring outside of their deployment (e.g., “near drowning of my son”). The remaining 68 participants recorded a traumatic event on the DTS that was either not a trauma as defined by the DSM-IV ( $n = 11$ ; e.g., “back pain,” “not being able to talk to my kids”), included multiple, discrete traumatic events in the narrative ( $n = 3$ ), was too vague to determine the nature of the trauma ( $n = 32$ ; e.g., “the way I was treated”), or reported no lifetime exposure to trauma ( $n = 21$ ). As a description of a bona fide traumatic event is necessary for the DTS to be considered valid (Davidson, 1996), these 68 participants were excluded. Thus, data from a total of 158 veterans were retained for analyses.

### 2.2. Measures

#### 2.2.1. Davidson Trauma Scale (Davidson, 1996; Davidson, Book, et al., 1997)

The DTS is a 17-item self-report questionnaire of posttraumatic stress symptoms, developed for use with trauma survivors. For the current study, the original Davidson, Book, et al. (1997) version was used. Respondents are first asked to record “the trauma that is most disturbing to you.” Next, respondents are asked to read each of the 17 items, and “consider how often in the last week the symptom troubled you and how severe it was.” The first five items specifically refer to reexperiencing or avoiding the disturbing event. The frequency and severity of the symptoms are recorded using 5-point, Likert-type scales. Frequency and severity scores were summed for each symptom, resulting in a total of 17 variables used in analyses (Elhai et al., 2006). The DTS total score was computed by adding all item responses together, with a possible range of 0–136. The three DTS subscales (reexperiencing, avoidance/numbing, and hyperarousal) and separate subscales for avoidance and for numbing (McDonald et al., 2008) were computed by adding all subscale items together and dividing by the total number used in the scale, resulting in a possible range of 0–5.

In an early validation study, the DTS demonstrated good internal consistency ( $\alpha = 0.99$ ), convergent validity (CAPS,  $R = 0.78$ ), divergent validity (extroversion,  $R = 0.04$ ), and concurrent validity, as well as strong test-retest reliability (Davidson, Book, et al., 1997; 0.86). A later study (Davidson, Tharwani, & Connor, 2002) demonstrated that the DTS is sensitive to treatment effects of selective serotonin reuptake inhibitors (SSRIs) for PTSD symptoms. Furthermore, the treatment effect size for the DTS was larger than the effect size for the Impact of Events Scale (Horowitz, Wilner, & Alvarez, 1979) and equal to those observed for the Clinician Administered PTSD Scale (CAPS; Blake et al., 1990) and Structured Interview for PTSD (SIP; Davidson, Malik, & Travers, 1997). Two studies have examined the factorial validity of the DTS.

Davidson, Book, et al. (1997) conducted exploratory factor analyses on data from 67 individuals with PTSD, and reported a six-factor structure that roughly corresponded to the three DSM-IV symptom clusters. More recently, using confirmatory factor analysis (CFA), McDonald et al. (2008) found that a four-factor structure for the DTS (reexperiencing, avoidance, numbing, and hyperarousal) was invariant across three veteran samples.

### 2.2.2. Structured Clinical Interview for DSM-IV-TR Axis I Disorders (First et al., 1994)

The SCID-I/P is a semi-structured interview used to diagnose DSM-IV Axis I disorders, and was used as the reference standard for a diagnosis of PTSD in this study. The SCID-I/P was administered by trained Masters or Doctorate-level research personnel. The reliability of interviewers on scoring a series of seven SCID training videos (full SCID I/P, mean Fleiss' kappa across 18 Axis I diagnoses) was excellent (Fleiss' kappa = 0.95).

### 2.2.3. Symptom Checklist-90-R (SCL-90-R; Derogatis, 1994)

The SCL-90-R is a 90-item self-report questionnaire used to evaluate a wide range of symptoms of psychopathology. Ratings are made on a 5-point scale. It has nine symptom scales and three global indices of distress. Internal consistency (alpha) for the subscales ranged from 0.77 (psychoticism) to 0.90 (depression) across two validation studies (Derogatis, 1994). The SCL-90-R subscales have demonstrated acceptable convergent and discriminant validity when compared to other multidimensional measures of psychopathology (e.g., MMPI-2, Middlesex Hospital Questionnaire) and when used to discriminate between anxiety and depressive disorders (Derogatis, 1994).

## 2.3. Analyses

Descriptive data for the sample, item- and scale-level characteristics (e.g., internal consistency), correlations, and mean group differences were computed using SPSS 15.0. Concurrent validity was examined by comparing the DTS total scores across the three veteran groups using analysis of variance (ANOVA). Pearson correlations between the DTS and the subscales of the SCL-90-R were conducted to examine convergent and divergent validity. In addition, standard multiple regression was employed to examine the unique relationship between SCL-90-R subscales and the DTS when all subscales are simultaneously modeled.

Receiver operating characteristic (ROC) curve analysis was used to examine the utility of the DTS compared to a reference standard of PTSD (i.e., SCID-I/P). The ROC curve, based upon signal detection theory, provides a visual and quantitative way of illustrating how well a continuous measure, or "test," properly classifies cases according to their known membership into one of two groups (Streiner, 2003). The accuracy of the test is graphically represented on an ROC curve, illustrating the change in sensitivity vs. (1 – specificity) as the cut-point on the test is varied. An ROC curve can be used to determine the optimal cut-point which maximizes true positives (i.e., sensitivity) while minimizing false positives (i.e., 1 – specificity) for a given application. The further the curve is positioned away from the diagonal (and toward high true positive rate and low false positive rate), the more accurate the test. An additional advantage of ROC analyses is that it provides a common metric, the "area under the curve" (AUC), with which to quantify the information value of assessment data that is independent of changes in the cutting-scores used for diagnostic decisions and prevalence rates (McFall & Treat, 1999). The AUC is a summary statistic that represents the ability of the variable to correctly classify cases into two groups for one or a range of scale values. More specifically, the AUC is the probability that two cases

randomly-drawn from the sample, one from each group, will both be classified correctly. A test that provides no information concerning group membership will have an AUC of 0.50, and is depicted on an ROC curve as a diagonal line. The AUC of a test increases with the accuracy of the test, to a maximum of 1.00.

Sensitivity (proportion of true positives), specificity (proportion of true negatives), positive predictive power (PPP; proportion of those testing positive who actually have the condition), negative predictive power (NPP; proportion of those testing negative who actually do not have the condition), and efficiency (proportion of individuals correctly classified) were calculated in addition to AUC statistics. As a diagnostic test is most efficient when 50% of the sample being tested has the condition of interest (Streiner, 2003), similar numbers of participants were included in each group to be compared to provide an optimal test of diagnostic efficiency. While sensitivity and specificity are considered fixed properties of a test (Streiner, 2003), the predictive power of a test is in part determined by the prevalence of a disorder in a given sample. Thus, PPP and NPP were computed using formulae provided by Streiner (2003) for several PTSD prevalence rates that approximate clinical conditions within the Department of Veterans Affairs health care system: 13% (Operations Enduring Freedom and Iraqi Freedom [OEF/OIF] veterans receiving care in DVA facilities; Seal, Bertenthal, Miner, Sen, & Marmar, 2007), 50% (OEF/OIF veterans with any mental health diagnosis; Seal et al., 2007), and 90% (treatment-seeking veterans at intake in a specialty PTSD clinic).

In addition to the standard scoring of the DTS, the utility of the symptom cluster method for testing diagnostic efficiency was also employed (e.g., Foa, Cashman, Jaycox, & Perry, 1997). The symptom cluster method is based on the DSM-IV criteria for PTSD, requiring the presence of one reexperiencing symptom, three avoidance/numbing symptoms, and two hyperarousal symptoms. Such a method may provide a more intuitive diagnostic procedure for those familiar with DSM-IV criteria. The present study is the first to test the utility of the symptom cluster method using the DTS.

## 3. Results

### 3.1. Sample characteristics

Gender, race, and other demographic data for the participants are presented in Table 1. Diagnostic interviews using the SCID-I/P revealed that 71 participants (45%) met criteria for an Axis I disorder, with about half of those (39, 55%) meeting criteria for PTSD. Of those with PTSD, 25 (64%) met criteria for a secondary Axis I disorder, consistent with the high levels of comorbidity described in other studies (Magruder et al., 2005). The most common diagnoses accompanying PTSD were Major Depressive Disorder (16, 41%) and Social Phobia (3, 8%). Of those meeting criteria for an Axis I disorder other than PTSD (alone or in combination with another Axis I disorder), the most common diagnoses were Major Depressive Disorder (13, 41%), Social Phobia (7, 22%), alcohol or substance abuse (7, 22%), and Anxiety, NOS (3, 9%). Eighty-seven participants (55%) did not meet criteria for any Axis I disorder. No one in the sample met criteria for a psychotic disorder.

### 3.2. Item-level and scale characteristics

Table 2 presents means, standard deviations, skew, kurtosis, corrected item-total correlations, and squared multiple correlations for items on the DTS. Internal consistency was good for the DTS total score (alpha = 0.97; Nunnally & Bernstein, 1994). Deletion of any one of the items would not substantially change the internal consistency of the DTS total score. However,

**Table 1**  
Participant demographics (n = 158)

	Number of participants	%
Median age	34 years	
Min–max age	21–64 years	
Males	126	80
Females	32	20
Marital status		
Married	80	50
Divorced/separated	27	18
Never married	45	29
Race/ethnicity		
Caucasian	68	44
African American	73	47
Hispanic	7	5
Other/unknown	10	4
Education		
High School/G.E.D.	80	51
Technical/trade	15	10
Associates degree	16	10
Bachelors degree	21	14
Post-bachelors degree	11	7
Other/unknown	13	8
Employment		
Full time	86	55
Part time	21	14
Not employed	49	31
Military rank		
E-1 to E-4	50	32
E-5 to E-8	93	59
Warrant officer	3	2
Officer	10	6
Unknown	2	1
Service in war zone	117	74.5 (per CES)
Combat exposure scale	M = 10.6 (10.4)	

evaluation of the corrected item total correlation (corrected  $r_{total}$ ) and the squared multiple correlation for each item suggested that item 7, “found yourself unable to recall important parts of the event” was a relatively poor fit to the scale.

Table 3 presents means, standard deviations, skew, kurtosis, Cronbach’s alphas, and Pearson correlations for the DTS total score, three subscale scores (reexperiencing, avoidance/numbing, and hyperarousal), and two separate subscale scores for avoidance and numbing items (McDonald et al., 2008). Internal consistency was good for the three conventional subscales (alpha = 0.92–0.94). Internal consistency of the subscale with the five numbing items (alpha = 0.89) and the subscale for the two avoidance items (alpha = 0.84) was adequate for group research, but was lower than acceptable for analysis of individual responses (Nunnally & Bernstein, 1994). The distribution of the DTS total score and subscale scores was generally positively skewed, with zero as the modal response. When only data from individuals with PTSD were assessed, skewness fell within normal limits for all (–0.12 to –0.43, S.E. = 0.38) but the arousal subscale (–0.80, S.E. = 0.38).

Two-way analyses of variance (ANOVA) were conducted to examine whether DTS total score differed for PTSD status (PTSD, no diagnosis) by gender (male, female) or race (White, Black). While there was a main effect of PTSD status ( $F[1,121] = 89.9, p < 0.001$ ), there was neither a main effect for gender ( $F[1,121] < 0.1, ns$ ), nor an interaction ( $F[1,121] = 0.8, ns$ ). Similarly for race, there was a main effect of PTSD status ( $F[1,108] = 127.3, p < 0.001$ ), but no main effect of race ( $F[1,108] < 0.1, ns$ ) or an interaction ( $F[1,108] = 0.1, ns$ ). For both gender and race, similar results were found with DTS subscale scores serving as the dependent variables.

### 3.3. Concurrent, convergent and divergent validity

Analyses of variance demonstrated diagnostic group differences on the DTS total score and subscales, supporting the concurrent validity of the DTS (Table 4). Participants diagnosed with PTSD ( $M = 79.6$ ) scored significantly higher than participants

**Table 2**  
Means, standard deviations, skew, kurtosis, corrected item-total correlations, and squared multiple correlations for items on the DTS

DTS items	DSM-IV symptom	Scale characteristics (n = 158)							Percentage of veterans endorsing symptom <sup>a</sup>		
		M	S.D.	Skew	Kurtosis	Corrected $r_{total}$	SMC	Alpha if deleted	No diagnosis (n = 32)	Other diagnosis (n = 32)	PTSD (n = 39)
1. Painful images, memories or thoughts of the event	B1	2.59	2.50	0.43	–1.13	0.77	0.73	0.97	38	69	92
2. Distressing dreams of the event	B2	1.95	2.60	0.90	–0.71	0.76	0.76	0.97	25	28	85
3. Felt as though the event was re-occurring	B3	1.66	2.40	1.05	–0.35	0.71	0.66	0.97	16	28	77
4. Upset by something which reminded you of the event	B4	2.26	2.39	0.61	–0.76	0.76	0.65	0.97	31	59	87
5. Avoiding any thoughts or feelings about the event	C1	2.30	2.87	0.76	–1.02	0.78	0.68	0.97	16	50	90
6. Avoiding doing things or going into situations which remind you about the event	C2	1.78	2.66	1.12	–0.28	0.81	0.73	0.97	19	34	74
7. Found yourself unable to recall important parts of the event	C3	0.93	2.14	2.29	4.09	0.56	0.37	0.97	6	13	44
8. Had difficulty enjoying things	C4	2.18	2.87	0.88	–0.79	0.80	0.76	0.97	13	53	80
9. Felt distant or cut off from other people	C5	2.49	3.04	0.68	–1.17	0.83	0.82	0.96	13	56	82
10. Been unable to have sad or loving feelings/generally felt numb	C6	1.89	2.71	1.10	–0.25	0.76	0.71	0.97	9	44	74
11. Found it hard to imagine having a long life span and fulfilling your goals	C7	1.82	2.78	1.16	–0.23	0.76	0.68	0.97	6	44	69
12. Had trouble falling asleep or staying asleep	D1	3.08	3.14	0.39	–1.43	0.82	0.74	0.97	31	69	95
13. Been irritable or had outbursts of anger	D2	2.21	2.75	0.78	–0.98	0.82	0.78	0.97	13	56	82
14. Had difficulty concentrating	D3	2.34	2.89	0.80	–0.88	0.85	0.80	0.96	13	53	92
15. Felt on edge, been easily distracted, or had to stay ‘on guard’	D4	2.54	3.03	0.66	–1.21	0.87	0.79	0.96	16	56	87
16. Been jumpy or easily startled	D5	2.25	2.97	0.88	–0.83	0.88	0.86	0.96	13	53	82
17. Been physically upset by reminders of the event	B5	1.58	2.51	1.30	0.28	0.78	0.75	0.97	9	34	74

Notes. Standard error for skew = 0.19. Standard error for kurtosis = 0.38. SMC, standard multiple correlation.

<sup>a</sup> Endorsement of each symptom (frequency >0) and experience of distress from that symptom (severity >0) within the past week.

**Table 3**

Means, standard deviations, skew, kurtosis, Cronbach's alphas, and inter-correlations for the DTS total and subscale scores ( $n = 158$ )

Scales	Scale characteristics						Pearson correlations					
	# Items	<i>M</i>	S.D.	Skew	Kurtosis	Alpha	1.	2.	3.	4.	5.	
1. Total score	17	35.52	37.57	0.86	-0.53	0.97	-					
2. Reexperiencing	5	2.01	2.16	0.89	-0.45	0.92	0.90	-				
3. Avoidance/numbing	7	2.04	2.56	0.96	-0.44	0.92	0.96	0.80	-			
4. Hyperarousal	5	1.86	2.27	1.04	-0.18	0.94	0.97	0.82	0.91	-		
5. Avoidance	2	2.48	2.65	0.70	-1.00	0.84	0.88	0.78	0.95	0.82	-	
6. Numbing	5	1.95	2.27	0.98	-0.28	0.89	0.93	0.73	0.93	0.90	0.77	-

Notes. Standard error for skew = 0.19. Standard error for kurtosis = 0.38. All correlations are significant at  $p < 0.001$ . Subscale means represent an average across items on that scale.

with no diagnosis ( $M = 14.7$ ) and those with a non-PTSD diagnosis ( $M = 37.6$ ). Similar findings were observed for the DTS subscales.

The DTS should demonstrate a modest relationship with measures of anxiety (i.e., convergent validity), as PTSD is considered an anxiety disorder. In contrast, the DTS should demonstrate a smaller relationship with measures of other symptoms, such as depression, psychosis, and hostility (i.e., divergent or discriminant validity). Table 5 displays Pearson correlations between the DTS and subscales from the SCL-90-R. Although correlations were moderate to strong between the DTS and all SCL-90-R subscales, correlations were strongest with two anxiety-related subscales (Obsessive-Compulsive and Anxiety), demonstrating convergent validity.

However, correlations between the DTS and other subscales were moderate to strong as well. To examine the independent ability of the subscales to predict the DTS, a standard multiple regression analysis was conducted with all SCL-90-R subscales placed into one step. The SCL-90-R subscales accounted for 62% of the variance in the DTS,  $F(9,139) = 27.8$ ,  $p < 0.001$ . Semi-partial correlations are presented in Table 5. One anxiety-related subscale, "Obsessive-Compulsive," was a significant predictor of DTS total scores after accounting for variance contributed by other subscales. The

"Anxiety" subscale accounted for the second-largest amount of variance in the DTS, although it was just beyond the threshold of significance ( $p = 0.06$ ). These findings suggest that the DTS shares a unique relationship with other measures of anxiety that it does not share with other measures of psychopathology (e.g., depression, psychoticism), providing evidence for divergent validity.

### 3.4. Diagnostic efficiency using the DTS total score

ROC curves and AUC statistics were generated for two comparisons. The first examined the ability of the DTS to discriminate between veterans with PTSD ( $n = 39$ ) and a random subsample of those with no Axis I diagnosis ( $n = 32$ ;  $M = 11.8$ , S.D. = 21.4). A second analysis was conducted to examine the ability of the DTS to discriminate between veterans with PTSD and those with other Axis I diagnoses ( $n = 32$ ).

As shown in Fig. 1, the DTS was excellent at discriminating between veterans with PTSD and those with no diagnosis (AUC = 0.95, S.E. = 0.03,  $p < 0.001$ , asymptotic 95% CI: lower bound = 0.89, upper bound = 1.00). A score of 32 provided the most efficient tradeoff of false positive to false negative classification (efficiency = 0.94; Table 6). Although its performance was still considered good by conventional standards, the DTS was less capable at discriminating between veterans with PTSD and those with mood or anxiety disorders but no PTSD (AUC = 0.83, S.E. = 0.05,  $p < 0.001$ , asymptotic 95% CI: lower bound = 0.74, upper bound = 0.92). The range of DTS scores that demonstrated comparable efficiency was present across a range of scores, with the highest efficiency at cut-points of 35 (efficiency = 0.77) and 75 (efficiency = 0.76; Table 6). Table 7 provides PPP and NPP values across a range of prevalence values for both comparisons.

### 3.5. Diagnostic efficiency using the DSM-IV symptom cluster method

The symptom cluster method provides an alternative to the cut-point method for evaluation diagnostic efficiency (Foa et al., 1997; Weathers, Ruscio, & Keane, 1999). This method requires endorse-

**Table 4**

Diagnostic group mean differences on the DTS total score and subscale scores

Scales	No diagnosis	Non- PTSD diagnosis	PTSD	ANOVA
<i>N</i>	86	32	39	
Total score				
<i>M</i>	14.7	37.6	79.6	$F(2,154) = 81.3$ , $p < 0.001$
S.D.	21.6	30.6	31.8	
Range	0–109	0–96	0–128	
Reexperiencing				
<i>M</i>	1.0 <sup>a</sup>	1.8 <sup>a</sup>	4.4	$F(2,155) = 58.7$ , $p < 0.001$
S.D.	1.5	1.6	1.9	
Avoidance/numbing				
<i>M</i>	0.8	2.0	4.5	$F(2,155) = 63.0$ , $p < 0.001$
S.D.	1.4	2.0	2.1	
Hyperarousal				
<i>M</i>	1.0	2.9	5.4	$F(2,155) = 68.5$ , $p < 0.001$
S.D.	1.6	2.5	2.1	
Avoidance				
<i>M</i>	0.9 <sup>b</sup>	2.0 <sup>b</sup>	4.7	$F(2,155) = 49.9$ , $p < 0.001$
S.D.	1.7	2.3	2.4	
Numbing				
<i>M</i>	0.7	2.1	4.2	$F(2,155) = 52.3$ , $p < 0.001$
S.D.	1.4	2.1	2.2	

Notes. Subscale means represent an average across items on that scale. Except where otherwise noted, means are significantly different using Bonferroni correction at  $p < 0.01$ .

<sup>a</sup>  $p < 0.10$ .

<sup>b</sup>  $p < 0.05$ .

**Table 5**

Pearson correlations of DTS total score with SCL-90-R subscale scores ( $n = 156$ )

Scale	Pearson correlation	Semi-partial correlation
Obsessive-Compulsive	0.77 <sup>***</sup>	0.15 <sup>**</sup>
Anxiety	0.73 <sup>***</sup>	0.10 <sup>†</sup>
Depression	0.69 <sup>***</sup>	0.05
Paranoid ideation	0.68 <sup>***</sup>	0.09 <sup>†</sup>
Somatization	0.66 <sup>***</sup>	0.07
Interpersonal sensitivity	0.64 <sup>***</sup>	-0.02
Phobic anxiety	0.62 <sup>***</sup>	0.02
Hostility	0.60 <sup>***</sup>	<±0.01
Psychoticism	0.60 <sup>***</sup>	<±0.01

<sup>\*\*</sup>  $p < 0.01$ .

<sup>\*\*\*</sup>  $p < 0.001$ .

<sup>†</sup>  $p < 0.10$ .

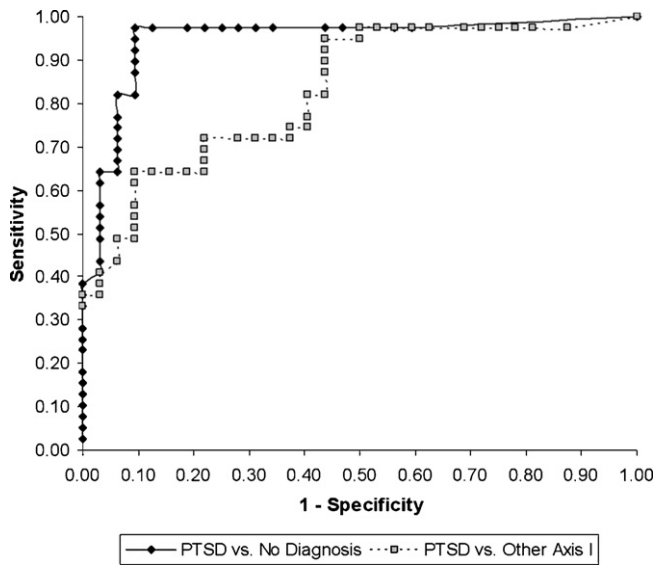


Fig. 1. Receiver operating characteristic (ROC) curve for various DTS total cut-point scores.

ment of one reexperiencing symptom, three avoidance/numbing symptoms, and two hyperarousal symptoms. Rates of endorsement for each symptom per diagnostic group are presented in Table 2. A symptom was scored as “present” if the respondent endorsed both the presence of the symptom (i.e., a “frequency” score of one or greater) and distress associated with the symptom (i.e., a “severity” score of one or greater). By this method, 36 of 39 veterans with PTSD were correctly classified (sensitivity = 0.92).

Table 6 Performance of the DTS in detecting PTSD at several cutoff scores

	24	32	35	40	48	75	95
DTS cutoff scores for PTSD (n = 39) vs. no diagnosis (n = 32)							
Counts							
True positive	38	38	37	34	32	25	15
False positive	6	6	3	3	2	1	0
False negative	1	1	2	5	7	14	24
True negative	26	26	29	29	30	31	32
Statistics							
Sensitivity	0.97	0.97	0.95	0.87	0.82	0.64	0.38
Specificity	0.81	0.91	0.91	0.91	0.94	0.97	1.00
PPP	0.86	0.93	0.93	0.95	0.94	0.96	1.00
NPP	0.96	0.97	0.94	0.85	0.81	0.69	0.57
LR+	5.20	10.39	10.12	9.30	13.13	20.51	na
LR-	31.69	35.34	17.67	7.07	5.22	2.70	1.63
Efficiency	0.90	0.94	0.93	0.89	0.87	0.79	0.66
DTS cutoff scores for PTSD (n = 39) vs. other Axis I diagnosis (n = 32)							
Counts							
True positive	38	38	37	34	32	25	15
False positive	19	16	14	14	13	3	1
False negative	1	1	2	5	7	14	24
True negative	13	16	18	18	19	29	31
Statistics							
Sensitivity	0.97	0.97	0.95	0.87	0.82	0.64	0.38
Specificity	0.41	0.50	0.56	0.56	0.59	0.91	0.97
PPP	0.67	0.70	0.73	0.71	0.71	0.89	0.94
NPP	0.93	0.94	0.90	0.78	0.73	0.67	0.56
LR+	1.64	1.95	2.17	1.99	2.02	6.84	12.31
LR-	15.84	19.50	10.97	4.39	3.31	2.52	1.57
Efficiency	0.72	0.76	0.77	0.73	0.72	0.76	0.65

Notes. PPP, positive predictive power; NPP, negative predictive power; LR+, likelihood ratio for a positive test; LR-, likelihood ratio for a negative test. LR+ is not available when Specificity = 1.00.

Table 7 Positive (PPP) and negative predictive power (NPP) at several DTS cutoff scores

	24	32	35	40	48	75	95
DTS cutoff scores for PTSD (n = 39) vs. no diagnosis (n = 32)							
13% prevalence							
PPP	0.44	0.61	0.60	0.58	0.66	0.75	1.00
NPP	1.00	1.00	0.99	0.98	0.97	0.95	0.92
50% prevalence							
PPP	0.84	0.91	0.91	0.90	0.93	0.95	1.00
NPP	0.97	0.97	0.95	0.88	0.84	0.73	0.65
90% prevalence							
PPP	0.98	0.99	0.99	0.99	0.99	0.99	1.00
NPP	0.78	0.80	0.66	0.44	0.37	0.23	0.15
DTS cutoff scores for PTSD (n = 39) vs. other Axis I diagnosis (n = 32)							
13% prevalence							
PPP	0.20	0.23	0.24	0.23	0.23	0.51	0.65
NPP	0.99	0.99	0.99	0.97	0.96	0.94	0.91
50% prevalence							
PPP	0.62	0.66	0.68	0.67	0.67	0.87	0.92
NPP	0.94	0.95	0.92	0.81	0.77	0.72	0.61
90% prevalence							
PPP	0.94	0.95	0.95	0.95	0.95	0.98	0.99
NPP	0.64	0.68	0.55	0.33	0.27	0.22	0.15

Twenty-nine of the 32 veterans with no diagnosis (specificity = 0.91, efficiency = 0.92) and only half of those with mood or anxiety disorders but not PTSD (16 of 32, specificity = 0.50, efficiency = 0.73) were correctly classified. To examine whether the test's efficiency could be improved by increasing the symptom severity criteria, the analysis was repeated, this time requiring symptoms to be rated at least “moderately distressing” (i.e., a “severity” score of two or greater). In this case, only 30 of 39 veterans with PTSD were correctly classified (sensitivity = 0.77). Thirty of 32 veterans with no diagnosis (specificity = 0.94, efficiency = 0.85) and 23 of 32 veterans with mood or anxiety disorders but not PTSD (specificity = 0.72, efficiency = 0.75) were correctly classified. Although increasing symptom severity criteria to “moderately distressing” resulted in improved specificity, sensitivity was attenuated. Thus, although these two symptom cluster methods demonstrated adequate efficiency in discriminating between veterans with PTSD and those with no Axis I diagnosis, they did not perform as well as the cut-score method.

#### 4. Discussion

The present study examined the validity and diagnostic efficiency of the DTS (Davidson, Book, et al., 1997) in a sample of veterans who have served since September 11th, 2001. Posttraumatic stress disorder is a relatively prevalent condition among these veterans, and there is a strong need for valid assessment tools. Brief self-report questionnaires, such as the DTS, can provide a considerable aid to the clinician and researcher in identifying those who are experiencing symptoms of PTSD. However, an invalid or misused diagnostic aid can cause more harm than good, considering the major impact that a diagnosis or positive screening can have in a patient's treatment and subsequently his or her quality of life (Streiner, 2003). Thus, it is critical that self-report measures of PTSD are valid in order to provide accurate aid to assessment, conduct quality research, and offer the best care to patients.

Results of the current study provide support for the DTS as a valid self-report measure of PTSD symptoms for veterans serving since 9/11, particularly in comparison to individuals without a psychiatric diagnosis. Concurrent validity was supported, as veterans with PTSD scored higher on the DTS than those with

other Axis I diagnoses or no diagnosis. Convergent and divergent validity were supported, as the DTS had a stronger relationship with anxiety-related symptom scales compared to those tapping other psychopathology. The DTS demonstrated good internal consistency ( $\alpha = 0.97$ ), matching or exceeding alpha reported for other PTSD symptom questionnaires (Norris & Hamblen, 2004). However, item 7, regarding the respondent's memory of the traumatic event, was endorsed by less than half of the participants with PTSD, and demonstrated relatively low inter-correlations with other DTS items. This finding is consistent with prior research indicating that loss of memory for the event is not a reliable predictor of PTSD (Davidson, Book, et al., 1997; Foa, Riggs, & Gershuny, 1995), and suggests that this item would benefit from rewording or perhaps closer scrutiny as a core symptom of PTSD.

Concerning diagnostic efficiency, potency of the DTS was dependent on the comparison group used for analyses: whereas diagnostic efficiency was excellent when discriminating between veterans with PTSD and veterans with no diagnosis ( $AUC = 0.95$ ), efficiency was attenuated, although still acceptable, when discriminating between PTSD and other Axis I diagnoses ( $AUC = 0.83$ ). When attempting to discriminate between veterans with PTSD and those with no Axis I diagnosis, the diagnostic efficiency of the DTS in the current study ( $AUC = 0.95$ ) was superior to that reported by Davidson, Book, et al. (1997;  $AUC = 0.88$ ). Whereas Davidson and colleagues reported that a DTS score of 40 provided the best efficiency (0.83), the current study found that a lower cut-point of 32 was more efficient in this sample of post-9/11 veterans (efficiency = 0.94).

Although the efficiency statistic is a common indicator of a measure's performance, the cut-point employed by a researcher or clinician should also be informed by the intended application. One should note that efficiency reflects the overall hit rate and thus places equal value on obtaining false positives and false negatives. In practice, other utility functions that place more or less emphasis on obtaining false positives may be preferred. For example, if the DTS was used as a PTSD screening tool to identify individuals who may benefit from more time and resource-intensive assessment, a clinician would likely want to use a lower cut-point to identify as many individuals with PTSD as possible (i.e., maximize sensitivity) while accepting the risk of increasing false positives. In another circumstance, a researcher with a limited budget may want to use a higher cut-point to reduce the number of false positives enrolled in the study (i.e., maximize specificity).

In addition to testing the ability of the DTS to discriminate between PTSD and healthy veterans, this study also examined its capability to discern PTSD from other Axis I disorders. This scenario is likely of more interest to clinicians and researchers in mental health clinics and psychiatric research, who may use PTSD symptom questionnaires to aid differential diagnosis. The DTS performed adequately in this circumstance ( $AUC = 0.83$ ), albeit not as well as when discriminating PTSD from healthy veterans. A cut-point of 35 provided the optimal balance of identifying as many veterans with PTSD as possible, while maximizing efficiency (0.77): 37 of 39 veterans with PTSD were correctly classified in this study, although specificity was poor, with 14 of 32 vets with other Axis I disorders mistakenly identified as having PTSD. If a clinician or researcher wishes to minimize false positives (while maximizing efficiency), a cut-point of 75 was best: in the current study, only three false positives (of 32, or 9%) were returned.

Results from our study examined the range of values for positive and predictive power that correspond to PTSD prevalence rates in several scenarios (post-9/11 veterans receiving care, post-9/11 veterans with any mental health diagnosis, 90% of treatment-seeking veterans in a specialty PTSD clinic), and these data illustrate that PPP and NPP vary depending on the prevalence of the

disorder in the clinic's population. When prevalence of a condition is low (e.g., primary care), a test is best used to rule out a condition but not to rule it in. For example, when applying the cutting score that maximized hit-rates between those with PTSD and without a psychiatric disorder in a primary care clinic setting, a positive test result would be correct 61% of the time. Thus, 39% of the time positive test results would be wrong. Similarly, when the prevalence of a condition is high such as a specialty PTSD clinic, a test is best used to rule in a condition but not to rule it out (Streiner, 2003).

Concerning the optimal scoring method for testing the diagnostic efficiency of the DTS, results of this study support the conventional cut-point method over a DSM-IV-based, symptom cluster method. Although the authors of the DTS utilized the cut-point method in an early validation study (Davidson, Book, et al., 1997), other PTSD symptom questionnaires have been scored using variations of both methods (e.g., Foa et al., 1997; Ruggiero, Del Ben, Scotti, & Rabalais, 2003). Results of the current study indicated that although the symptom cluster method was effective at correctly classifying veterans with PTSD from those with no Axis I disorders, the cut-point method was more efficient. These findings are consistent with a recent review that found little benefit in utilizing more complex scoring methods instead of the conventional cut-point strategy (Brewin, 2005).

An important issue concerning PTSD symptom questionnaires that needs further research is the potential impact of anchoring responses to one particular traumatic event (Norris & Hamblen, 2004). Clarifying the nature of the disturbing event would appear especially important, given evidence that persons with other psychiatric disturbances, such as major depression, can report symptoms consistent with PTSD even though they have not been exposed to a Criterion A traumatic event (Bodkin, Pope, Detke, & Hudson, 2007). The DTS attempts to capture the index trauma by requiring respondents to write a brief description of the Criterion A traumatic event "that is most disturbing to you" (Davidson, 1996). For this reason, the DTS ostensibly has an advantage over other PTSD measures that do not document the index trauma, such as the PCL-C (Blanchard, Jones-Alexander, Buckley, & Forneris, 1996) and the PC-PTSD (Prins et al., 2003) when used as a PTSD screening tool. However, in this study, 20% of the participants recorded a disturbing event on the DTS that could not be clearly identified as a discrete DSM-IV PTSD Criterion A1 "trauma." Certainly, many of the participants had a bona fide trauma in mind, as evinced by the 16 of the 47 excluded participants who met SCID I/P diagnosis for PTSD. Perhaps some participants refrained from providing detailed trauma descriptions due to fearful avoidance of reminders, or other motivational factors such as guilt, shame, or fatigue. For whatever reasons, these cases demonstrate the limitation of using an open-ended response format to record traumatic stressors on self-report questionnaires. Although rarely done in practice and not generally reported in previous studies using self-report PTSD instruments, it would seem important to assess whether the event meets criteria for a "traumatic event" prior to the respondent completing the symptom inventory. In this regard, it may be helpful to anchor the DTS symptom ratings to the outcome of a trauma rating scale, such as the Traumatic Life Events Questionnaire (Kubany et al., 2000; cf. PDS, Foa et al., 1997).

It is notable that this sample entirely consisted of veterans who have served since September 9th, 2001. Although this specificity will be useful to those working with veterans who have served in the post-9/11 era, the extent to which these results generalize to other populations is unclear. In addition, although the participant group employed for this study was diverse, the sample was not large enough to fully examine ethnic or racial differences in diagnostic efficiency for the DTS. Further research is needed to



determine generalizability of these findings across demographic groups (e.g., race, gender, and first language), the range of symptom severity and comorbidity, and a variety of settings (i.e., primary care).

Our results suggest that use of the DTS as a diagnostic screening in settings where rates of other Axis I disorders are high (e.g., a mental health clinic), particularly at the level of usual recommended screening cutoff scores, substantial diagnostic errors could result. Several symptoms associated with PTSD are also common in other Axis I mental disorders (e.g., concentration problems in Major Depressive Disorder). Thus, it is not surprising that those with an Axis I diagnosis other than PTSD had elevated DTS scores relative to those with no diagnosis, and subsequently were far more difficult to distinguish from those with PTSD. This result is also consistent with other research suggesting that PTSD symptom instruments overlap substantially with other psychological distress symptoms (Lauterbach, Vrana, King, & King, 1997). Results are also consistent with a taxometric approach to psychopathology which acknowledges the nonspecific symptoms of distress reported by patients with PTSD, depressive disorders, and Generalized Anxiety Disorder (e.g., Bodkin et al., 2007; O'Donnell, Creamer, & Pattison, 2004; Watson, Gamez, & Simms, 2005). In addition, these findings provide practical support for the view that only symptoms with diagnostic specificity be retained in future revisions of the DSM (McHugh & Treisman, 2007; Spitzer, First, & Wakefield, 2007).

Unfortunately, no reports are available concerning the ability of the PTSD Checklist (PCL) or the Posttraumatic Diagnostic Scale (Foa et al., 1997) to discriminate between PTSD and other Axis I disorders. Our results suggest that this would be an important avenue for future studies evaluating self-report instruments for PTSD. In addition, we urge clinicians and researchers who utilize a measure's diagnostic efficiency statistics as part of their diagnostic assessment to consider whether the comparison group used in the validation study is an appropriate match to the population they serve. Furthermore, we reiterate the consensus of the authors of the DTS that it should not be used alone to make a diagnosis of PTSD. Instead, the DTS is foremost a measure of symptom severity, and secondly an ally in comprehensive clinical assessment.

## Acknowledgements

Preparation of this manuscript was in part supported by Office of Research and Development Clinical Science, Department of Veterans Affairs, K24DA016388, K23MH073091, 2R01CA081595, and R21DA019704 and HL54780. The views expressed in this presentation are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the National Institutes of Health. We would like to acknowledge the contributions of the OIF/OEF Registry Workgroup (Ms. Kimberly Green and Drs. Larry Tupler, Ryan Wagner and Christine Marx), and principal investigators for the OEF/OIF Registry at sites of data collection other than the Durham VAMC (Drs. Marinell Miller-Mumford, Antony Fernandez, Katherine Taber, and Ruth Yoash-Gantz).

## References

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2004). *Practice Guideline for the Treatment of Patients with Acute Stress Disorder and Post-traumatic Stress Disorder*. Retrieved May 2006.

Andrykowski, M. A., Cordova, M. J., Studts, J. L., & Miller, T. W. (1998). Posttraumatic stress disorder after treatment for breast cancer: prevalence of diagnosis and use of the PTSD Checklist-Civilian Version (PCL-C) as a screening instrument. *Journal of Consulting and Clinical Psychology, 66*, 586–590.

Beckham, J. C., Moore, S. D., Feldman, M. E., Hertzberg, M. A., Kirby, A. C., & Fairbank, J. A. (1998). Health status, somatization, and severity of posttraumatic stress disorder in Vietnam combat veterans with posttraumatic stress disorder. *American Journal of Psychiatry, 155*, 1565–1569.

Bisson, J., & Andrew, M. (2005). Psychological treatment of post-traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews 2: Art. No.: CD003388*.

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

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

Bodkin, J. A., Pope, H. G., Detke, M. J., & Hudson, J. I. (2007). Is PTSD caused by traumatic stress? *Journal of Anxiety Disorders, 21*, 176–182.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., et al. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of Internal Medicine, 138*, W1–W12.

Brewin, C. R. (2005). Systematic review of screening instruments for adults at risk of PTSD. *Journal of Traumatic Stress, 18*, 53–62.

Cook, J. M., Elhai, J. D., & Areán, P. A. (2005). Psychometric properties of the PTSD Checklist with older primary care patients. *Journal of Traumatic Stress, 18*, 371–376.

Coyne, J. C., & Thompson, R. (2007). Posttraumatic stress syndromes: useful or negative heuristics? *Journal of Anxiety Disorders, 21*, 223–229.

Davidson, J. (1996). *Davidson Trauma Scale (DTS)*. Tonawanda, New York: Multi-Health Systems Inc.

Davidson, J., Tharwani, H. M., & Connor, K. M. (2002). Davidson Trauma Scale (DTS): normative scores in the general population and effect sizes in placebo-controlled SSRI trials. *Depression & Anxiety, 15*, 75–78.

Davidson, J. R. T., Book, S. W., Colket, J. T., Tupler, L. A., Roth, S., David, D., et al. (1997). Assessment of a new self-rating scale for post-traumatic stress disorder. *Psychological Medicine, 27*, 153–160.

Davidson, J. R. T., Malik, M. A., & Travers, J. (1997). Structured interview for PTSD (SIP): psychometric validation for DSM-IV criteria. *Depression & Anxiety (1091–4269), 5*, 127–129.

Department of Veterans Affairs & Department of Defense. (2004). *VA/DoD Clinical Practice Guideline for the Management of Post-Traumatic Stress*. Washington, DC: Veterans Health Administration, Department of Veterans Affairs and Health Affairs, Department of Defense.

Derogatis, L. R. (1994). *SCL-90-R: administration, scoring, and procedures manual* (3rd ed.). Minneapolis, MN: NCS Pearson Inc.

Dobie, D. J., Kivlahan, D. R., Maynard, C., Bush, K. R., McFall, M., Epler, A. J., et al. (2002). Screening for post-traumatic stress disorder in female Veteran's Affairs patients: validation of the PTSD checklist. *General Hospital Psychiatry, 24*, 367–374.

Dohrenwend, B. P., Turner, J. B., Turse, N. A., Adams, B. G., Koenen, K. C., & Marshall, R. (2007). Continuing controversy over the psychological risks of Vietnam for U.S. veterans. *Journal of Traumatic Stress, 20*, 449–465.

Elhai, J. D., Lindsay, B. M., Gray, M. J., Grubaugh, A. L., North, T. C., & Frueh, B. C. (2006). Examining the uniqueness of frequency and intensity symptom ratings in post-traumatic stress disorder assessment. *Journal of Nervous and Mental Disease, 194*, 940–944.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1994). *Structured clinical interview for DSM-IV axis I disorders—patient edition (SCID-I/P, Version 2.0)*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: the Posttraumatic Diagnostic Scale. *Psychological Assessment, 9*, 445–451.

Foa, E. B., Riggs, D. S., & Gershuny, B. S. (1995). Arousal, numbing, and intrusion: symptom structure of PTSD following assault. *American Journal of Psychiatry, 152*, 116–120.

Forbes, D., Creamer, M., & Biddle, D. (2001). The validity of the PTSD checklist as a measure of symptomatic change in combat-related PTSD. *Behaviour Research and Therapy, 39*, 977–986.

Hankin, C. S., Spiro, A., III, Miller, D. R., & Kazis, L. (1999). Mental disorders and mental health treatment among U.S. Department of Veterans Affairs outpatients: the Veterans Health Study. *The American Journal of Psychiatry, 156*, 1924–1930.

Hoge, C. W., Auchterlonie, J. L., & Milliken, C. S. (2006). Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. *Journal of the American Medical Association, 295*, 1023–1032.

Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine, 351*, 13–22.

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

Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment, 17*, 251–255.

Jordan, B. K., Marmar, C. R., Fairbank, J. A., Schlenger, W. E., Kulka, R. A., Hough, R. L., et al. (1992). Problems in families of male Vietnam veterans with Posttraumatic Stress Disorder. *Journal of Consulting & Clinical Psychology, 60*, 916–926.

Kubany, E. S., Leisen, M. B., Kaplan, A. S., Watson, S. B., Haynes, S. N., Burns, K., et al. (2000). Development and preliminary validation of a brief broad-spectrum measure of trauma exposure: the traumatic life events questionnaire. *Psychological Assessment, 12*, 210–224.

- Lang, A. J., Laffaye, C., Satz, L. E., Dresselhaus, T. R., & Stein, M. B. (2003). Sensitivity and specificity of the PTSD Checklist in detecting PTSD in female veterans in primary care. *Journal of Traumatic Stress, 16*, 257.
- Lauterbach, D., Vrana, S., King, D. W., & King, L. A. (1997). Psychometric properties of the Civilian Version of the Mississippi PTSD scale. *Journal of Traumatic Stress, 10*, 499–513.
- Magruder, K. M., Frueh, B. C., Knapp, R. G., Davis, L., Hamner, M. B., Martin, R. H., et al. (2005). Prevalence of posttraumatic stress disorder in Veterans Affairs primary care clinics. *General Hospital Psychiatry, 27*, 167–179.
- McDonald, S. D., Beckham, J. C., Morey, R., Marx, C., Tupler, L. A., & Calhoun, P. S. (2008). Factorial invariance of Posttraumatic Stress Disorder symptoms across three veteran samples. *Journal of Traumatic Stress, 21*(3), 309–317.
- McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50*, 215–241.
- McHugh, P. R., & Treisman, G. (2007). PTSD: a problematic diagnostic category. *Journal of Anxiety Disorders, 21*, 211–222.
- Norris, F. H., & Hamblen, J. L. (2004). Standardized self-report measures of civilian trauma and PTSD. In: J. Wilson & T. M. Keane (Eds.), *Assessing psychological, trauma and PTSD* (2nd ed., pp. 63–102). New York: Guilford Press.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill Inc.
- O'Donnell, M. L., Creamer, M., & Pattison, P. (2004). Posttraumatic Stress Disorder and depression following trauma: understanding comorbidity. *The American Journal of Psychiatry, 161*, 1390–1396.
- Prins, A., Ouimette, P., Kimerling, R., Cameron, R. P., Hugelshofer, D. S., Shaw-Hegwer, J., et al. (2003). The primary care PTSD screen (PC-PTSD): development and operating characteristics. *Primary Care Psychiatry, 9*, 9–14.
- Ruggiero, K. J., Del Ben, K., Scotti, J. R., & Rabalais, A. E. (2003). Psychometric properties of the PTSD Checklist—civilian version. *Journal of Traumatic Stress, 16*, 495–502.
- Seal, K. H., Bertenthal, D., Miner, C. R., Sen, S., & Marmar, C. (2007). Bringing the war back home: mental health disorders among 103,788 US veterans returning from Iraq and Afghanistan seen at Department of Veterans Affairs facilities. *Archives of Internal Medicine, 167*, 476–482.
- Shalev, A. Y., Freedman, S., Peri, T., Brandes, D., Sahar, T., Orr, S. P., et al. (1998). Prospective study of posttraumatic stress disorder and depression following trauma. *American Journal of Psychiatry, 155*, 630–637.
- Spitzer, R. L., First, M. B., & Wakefield, J. C. (2007). Saving PTSD from itself in DSM-V. *Journal of Anxiety Disorders, 21*, 233–241.
- Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). *Structured Clinical Interview for DSM-III-R (SCID)*. New York: Biometrics Research Department, New York State Psychiatric Institute.
- Streiner, D. L. (2003). Diagnosing tests: using and misusing diagnostic and screening tests. *Journal of Personality Assessment, 81*, 209–219.
- Taft, C. T., Stern, A. S., King, L. A., & King, D. W. (1999). Modeling physical health and functional health status: the role of combat exposure, posttraumatic stress disorder, and personal resource attributes. *Journal of Traumatic Stress, 12*, 3–23.
- Walker, E. A., Katon, W., Russo, J., Ciechanowski, P., Newman, E., & Wagner, A. W. (2003). Health care costs associated with posttraumatic stress disorder symptoms in women. *Archives of General Psychiatry, 60*, 369–374.
- Watson, D., Gamez, W., & Simms, L. J. (2005). Basic dimensions of temperament and their relation to anxiety and depression: a symptom-based perspective. *Journal of Research in Personality, 39*, 46–66.
- Weathers, F. W., Ruscio, A. M., & Keane, T. M. (1999). Psychometric properties of nine scoring rules for the Clinician-Administered Posttraumatic Stress Disorder Scale. *Psychological Assessment, 11*, 124–133.
- Wohlfarth, T. D., van den Brink, W., Winkel, F. W., & ter Smitten, M. (2003). Screening for posttraumatic stress disorder: an evaluation of two self-report scales among crime victims. *Psychological Assessment, 15*, 101–109.