

The Clinician-Administered PTSD Scale for *DSM–5* (CAPS-5): Development and Initial Psychometric Evaluation in Military Veterans

Frank W. Weathers
Auburn University

Michelle J. Bovin
National Center for PTSD at VA Boston Healthcare System,
Boston, Massachusetts, and Boston University School of Medicine

Daniel J. Lee
Auburn University, and VA Boston Healthcare System,
Boston, Massachusetts

Denise M. Sloan
National Center for PTSD at VA Boston Healthcare System,
Boston, Massachusetts, and Boston University School of Medicine

Paula P. Schnurr
National Center for PTSD, White River Junction, Vermont, and
Geisel School of Medicine at Dartmouth

Danny G. Kaloupek, Terence M. Keane, and
Brian P. Marx
National Center for PTSD at VA Boston Healthcare System,
Boston, Massachusetts, and Boston University School of Medicine

The Clinician-Administered PTSD Scale (CAPS) is an extensively validated and widely used structured diagnostic interview for posttraumatic stress disorder (PTSD). The CAPS was recently revised to correspond with PTSD criteria in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM–5)*; American Psychiatric Association, 2013). This article describes the development of the CAPS for *DSM–5* (CAPS-5) and presents the results of an initial psychometric evaluation of CAPS-5 scores in 2 samples of military veterans ($Ns = 165$ and 207). CAPS-5 diagnosis demonstrated strong interrater reliability ($\kappa = .78$ to 1.00 , depending on the scoring rule) and test–retest reliability ($\kappa = .83$), as well as strong correspondence with a diagnosis based on the CAPS for *DSM–IV* (CAPS-IV; $\kappa = .84$ when optimally calibrated). CAPS-5 total severity score demonstrated high internal consistency ($\alpha = .88$) and interrater reliability ($ICC = .91$) and good test–retest reliability ($ICC = .78$). It also demonstrated good convergent validity with total severity score on the CAPS-IV ($r = .83$) and PTSD Checklist for *DSM–5* ($r = .66$) and good discriminant validity with measures of anxiety, depression, somatization, functional impairment, psychopathy, and alcohol abuse ($rs = .02$ to $.54$). Overall, these results indicate that the CAPS-5 is a psychometrically sound measure of *DSM–5* PTSD diagnosis and symptom severity. Importantly, the CAPS-5 strongly corresponds with the CAPS-IV, which suggests that backward compatibility with the CAPS-IV was maintained and that the CAPS-5 provides continuity in evidence-based assessment of PTSD in the transition from *DSM–IV* to *DSM–5* criteria.

Public Significance Statement

This study evaluated the *DSM–5* version of the Clinician-Administered PTSD Scale (CAPS-5), a widely used structured interview for posttraumatic stress disorder, in 2 samples of military veterans. Results indicated that the CAPS-5 is psychometrically sound and corresponds closely with the previous *DSM–IV* version of the CAPS.

Keywords: CAPS-5, PTSD, *DSM–5*, structured interview, psychometric

This article was published Online First May 11, 2017.

Frank W. Weathers, Department of Psychology, Auburn University; Michelle J. Bovin, National Center for PTSD, VA Boston Healthcare System, Boston, Massachusetts, and Department of Psychiatry, Boston University School of Medicine; Daniel J. Lee, Department of Psychology, Auburn University, and VA Boston Healthcare System, Boston, Massachusetts; Denise M. Sloan, National Center for PTSD, VA Boston Healthcare System, and Department of Psychiatry, Boston University School of Medicine; Paula P. Schnurr, National Center for PTSD, VA Medical Center, White River Junction, Vermont, and Department of Psychiatry, Geisel School of Medicine at Dartmouth; Danny G. Kaloupek, Terence M.

Keane, and Brian P. Marx, National Center for PTSD, VA Boston Healthcare System, and Department of Psychiatry, Boston University School of Medicine.

A portion of the current study was funded by Department of Veteran Affairs Merit Grant (I01 CX000467) awarded to Denise M. Sloan.

Frank W. Weathers, Michelle J. Bovin, and Daniel J. Lee receive consultant fees for providing CAPS-5 training to multiple research projects.

Correspondence concerning this article should be addressed to Frank W. Weathers, Department of Psychology, Auburn University, 226 Thach Hall, Auburn, AL 36849. E-mail: weathfw@auburn.edu

The Clinician-Administered PTSD Scale (CAPS) is a structured diagnostic interview that assesses posttraumatic stress disorder (PTSD) diagnostic status and symptom severity. Developed in 1989 at the National Center for PTSD (Blake et al., 1990), the CAPS has been extensively validated (Weathers, Keane, & Davidson, 2001); widely used in clinical, research, and forensic settings (Elhai, Gray, Kashdan, & Franklin, 2005); and generally recognized in the field of traumatic stress as a benchmark criterion measure of PTSD. Notable features of the CAPS include (a) assessment of all PTSD criteria plus associated features such as dissociation; (b) global ratings of distress, impairment, response validity, symptom severity, and improvement since a previous assessment; (c) both dichotomous (present/absent) and continuous ratings for individual symptoms and overall disorder; (d) separate assessment of symptom frequency and intensity; (e) behaviorally anchored prompts and rating scales; and (f) assessment of trauma-relatedness for individual symptoms not inherently linked to the trauma (e.g., loss of interest, estrangement, difficulty concentrating).

The CAPS was recently updated to reflect changes to the PTSD criteria for the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013)*. The revision process involved examination of the relevant empirical literature, consideration of qualitative critiques from CAPS users over the 20 years since the previous revision for *DSM-IV*, and numerous discussions among CAPS authors and their colleagues. In this article, we describe the specific revisions to the CAPS for *DSM-5* and present the results of an initial psychometric evaluation of CAPS-5 scores in two samples of military veterans.

Goals for the CAPS Revision

Maintain DSM Correspondence

There were three goals in revising the CAPS for *DSM-5*. The first goal was to accurately reflect *DSM-5* PTSD criteria. Since its inception, the CAPS has been a *DSM*-correspondent instrument, and the CAPS-5 continues this tradition. Thus, paralleling the changes to the *DSM-5* PTSD criteria, the CAPS revision involved (a) elimination of Criterion A2, (b) addition of three new symptoms and reconceptualization of several existing symptoms, and (c) separation of the avoidance and numbing symptom cluster into two clusters labeled avoidance and negative alterations in cognition and mood (NACM; see Weathers, Marx, Friedman, & Schnurr, 2014).

Streamline Administration and Scoring

The second goal for the CAPS revision was to streamline administration and scoring. This was accomplished by standardizing and simplifying conversion of symptom frequency and intensity ratings into symptom severity ratings and dichotomous scores. In addition, prompts were reorganized in a top-to-bottom format following a standard sequence across symptoms, namely, an initial prompt describing the symptom, followed by a prompt for examples and then intensity, frequency, and trauma-related prompts. Interviewers can quickly achieve proficiency with CAPS-5 admin-

istration simply by starting with the initial prompt for a symptom and working down the page to the last prompt.

Limitations of scoring on previous CAPS. Regarding scoring, there were two key limitations to prior versions of the CAPS. First, for some applications frequency and intensity ratings were summed to create a 9-point (0 to 8) symptom severity scale. Conceptually this approach gives equal weight to symptom frequency and intensity, which may not always be appropriate because the same severity score could result from either a high frequency but low intensity behavior (not likely to be considered a symptom) or a low frequency and high intensity behavior (likely to be considered a symptom).

Second, obtaining a dichotomous symptom score required some method for converting 0 to 4 frequency and intensity ratings into present/absent ratings. This led to the development of a number of rationally and empirically derived scoring rules, ranging from lenient to stringent (Weathers, Ruscio, & Keane, 1999). The variety of feasible rules underscored the ambiguity involved in converting continuous scores into dichotomous scores and illustrated the need for clinicians and investigators to explicitly justify their choice of a rule for a given assessment task. Nonetheless, some of the CAPS scoring rules were too complex for interviewers to apply in real time while administering the interview. Consequently, the rationally derived F1/I2 rule (whereby a symptom is considered present if Frequency is rated 1 or higher and Intensity is rated 2 or higher) became the default, in part because it represents a conceptual minimum threshold, but also because it is easy to apply. However, F1/I2 is the most lenient of the established CAPS scoring rules, and may be too lenient in that it can result in an individual receiving a PTSD diagnosis but having a total severity score in the subthreshold or even asymptomatic range.

Simplification of scoring for CAPS-5. Taking these issues into consideration, item scoring was substantially simplified for the CAPS-5 in several ways. First, frequency and intensity continue to be assessed and rated separately. Intensity is rated as *minimal*, *clearly present*, *pronounced*, and *extreme*, and frequency is recorded directly as reported by the respondent, either as a number of times or a percentage of time, depending on the symptom. Second, drawing on the F1/I2 rule as well as two empirically validated, clinician-rated scoring rules from research conducted in the mid-1990s (i.e., CR60 and CR75; see Weathers et al., 1999), a scoring system—based on both rational and empirical considerations—was developed for converting frequency and intensity information into a single 5-point (0 to 4) symptom severity scale. The anchor points for this severity scale are 0 = *absent*, 1 = *mild/subthreshold*, 2 = *moderate/threshold*, 3 = *severe/markedly elevated*, and 4 = *extreme/incapacitating*.

The frequency and intensity thresholds for the two key severity ratings (2 = *moderate/threshold* and 3 = *severe/markedly elevated*) are provided in the interview form for each symptom so that interviewers can refer directly to them to make the appropriate severity rating. A severity rating of 2 generally requires a minimum frequency of at least twice a month or some of the time (20% to 30%) and a minimum intensity of *clearly present*. A severity rating of 3 generally requires a minimum frequency of twice a week and a minimum intensity of *pronounced*. Finally, a symptom is considered present and subsequently counted toward a PTSD diagnosis if its severity rating is 2 or higher—this is the SEV2 rule, the basic CAPS-5 symptom scoring rule. Conceptually, SEV2 is

slightly more stringent than F1/I2 in that the minimum frequency for SEV2 is twice a month, whereas the minimum frequency for F1/I2 is once or twice a month.

Maintain Backward Compatibility

The third goal in developing the CAPS-5 was to maintain a high level of backward compatibility with previous versions of the CAPS. This would provide continuity in the use of trademark CAPS features, which have good content validity and clinical utility, and facilitate integration of new findings based on the CAPS-5 with the extensive existing literature on the CAPS. Most prompts and rating scale anchors were retained verbatim or slightly rephrased for clarity based on user feedback. In addition, other features of the CAPS-5 were retained from previous versions and revised to reflect *DSM-5* criteria, including (a) the Life Events Checklist, used in conjunction with the CAPS-5 Criterion A assessment section to identify an index event for symptom inquiry; (b) the trauma-related inquiry, used for symptoms not inherently linked to the index event; (c) items assessing global distress and functional impairment, response validity, overall severity, and improvement since a previous assessment; and (d) items assessing depersonalization and derealization, now used to assess the new dissociative subtype of PTSD.

Revision Process

All revisions for the CAPS-5 were drafted by the first author, in close consultation with CAPS-5 coauthors. New CAPS-5 items to assess new *DSM-5* symptoms were written in the style of existing CAPS items and closely followed *DSM-5* criterion language. All revisions were reviewed by numerous experts in PTSD assessment, including the CAPS-5 authors, colleagues at the National Center for PTSD, and the chair of and advisors to the Trauma/Stress-Related and Dissociative Disorders Sub-Work Group (Friedman, 2013). The revision process addressed key aspects of content validity (Haynes, Richard, & Kubany, 1995)—including item content, rating scale format, and instructions for standard administration and scoring—and involved circulating drafts among the authors and other trauma experts until consensus was reached regarding the final form of the interview.

The Present Study

The research described in the following text was an initial psychometric evaluation of CAPS-5 scores involving two samples of military veterans. Although this is the first comprehensive evaluation of the CAPS-5, other studies have presented limited psychometric evidence. For example, Marmar et al. (2015) used the CAPS-5 as a primary diagnostic measure in the National Vietnam Veterans Longitudinal Study. They found excellent interrater reliability, with a kappa of .93, based on independent, blinded ratings of audio-recorded CAPS-5 interviews. They also found excellent correspondence in signal detection analyses between the CAPS-5 and the PCL-5, PCL for *DSM-IV*, and the Mississippi Scale for Combat-Related PTSD (Keane, Caddell, & Taylor, 1988), providing strong evidence of convergent validity.

In addition, Foa et al. (2016) used the CAPS-5 in their evaluation of the PTSD Symptom Scale Interview for *DSM-5* (PSSI-5).

They found good convergent validity between PSSI-5 total score and CAPS-5 total score, with correlation of .72. However, using the CAPS-5 as the criterion they found only moderate correspondence between PSSI-5 and CAPS-5 diagnostic status, with a sensitivity of .82, specificity of .71, and kappa of .49. We address the issue of correspondence between the CAPS-5 and PSSI-5 more fully in the discussion.

The first three phases of the present study all involved Sample 1. In Phase 1 different clinicians working independently administered both the CAPS-5 and the CAPS for *DSM-IV* (CAPS-IV) in counterbalanced order. In Phase 2 different clinicians administered the CAPS-5 two separate times. This design is commonly referred to as test-retest reliability, but because it involves different clinicians it combines test-retest (occasions) and alternate form (interviewers) methods. Thus, the resulting reliability estimate reflects two sources of error and is technically a coefficient of stability and interrater equivalence (Crocker & Algina, 1986). Phase 3 respondents were administered a single CAPS-5. All Sample 1 participants also completed a battery of questionnaires.

Psychometric properties of CAPS-5 scores evaluated in the data from Sample 1 included internal consistency (alpha coefficients and interitem correlations), interrater and test-retest reliability (intraclass correlations for continuous symptom severity scores and kappa coefficients for dichotomous diagnosis), and convergent and discriminant validity. Sample 2 participants were administered a single CAPS-5. We combined data from Samples 1 and 2 to conduct a confirmatory factor analysis (CFA) of CAPS-5 scores.

We hypothesized that CAPS-5 scores would demonstrate high internal consistency, interrater reliability, and test-retest reliability; strong correspondence with CAPS-IV scores; and good convergent and discriminant validity with scores on various questionnaire measures of PTSD and other relevant constructs. Further, we hypothesized that CAPS-5 and CAPS-IV scores would demonstrate a similar pattern of associations with measures of PTSD and other constructs, with both versions of the CAPS correlating (a) strongly with other measures of PTSD; (b) moderately with measures of constructs closely related to PTSD, including depression, anxiety, somatization, and functional impairment; and (c) weakly with measures of antisocial personality and alcohol abuse.

Regarding PTSD diagnosis, we hypothesized that there would be moderate to strong correspondence between the CAPS-5 using SEV2 and the CAPS-IV using F1/I2, but that the strongest correspondence between CAPS-5 and CAPS-IV PTSD diagnosis would be found using scoring rules that required a minimum total severity score in addition to the requisite symptoms. The possibility of only moderate correspondence for CAPS-5/SEV2 and CAPS-IV/F1-I2 was based on the facts that (a) both the PTSD criteria and CAPS format were revised for *DSM-5*, thereby creating potential diagnostic discordance; and (b) F1/I2 is the most lenient CAPS-IV scoring rule, and CAPS-5 SEV2 is conceptually only slightly less lenient, which means that some individuals could be just at a conceptual diagnostic threshold (i.e., meet symptom criteria but have a low total severity score), possibly resulting in greater instability in diagnostic status across repeated measurements (i.e., interrater or test-retest reliability). Thus, to fully evaluate backward compatibility we sought to calibrate the CAPS-5 with the CAPS-IV by identifying the scoring rules on each that optimize their correspondence on PTSD diagnosis. Finally, we hypothesized that the four-factor *DSM-5* model of PTSD would provide ade-

quate fit in CFA, but that recently proposed six- and seven-factor models (Armour et al., 2015) would provide superior fit.

Method

Participants. Sample 1 consisted of 167 veterans recruited at a VA Healthcare System for a study designed to validate both the PTSD Checklist for *DSM-5* (PCL-5; Weathers, et al., 2013) and the CAPS-5 (Weathers, Blake, et al., 2013). A subset of these data was used previously to validate the PCL-5 (Bovin et al., 2016). This study followed the second version of the Quality Assessment of Diagnostic Accuracy Studies guidelines (QUADAS-2; Whiting et al., 2011), which minimizes the influence of various sources of bias that can affect diagnostic utility studies (see Bovin et al., 2016). This study was open to all veterans who were aged 18 or older who could read written materials in English. Potential participants were screened for trauma exposure and PTSD symptoms with the Brief Trauma Questionnaire (Schnurr et al., 2002) and Primary Care PTSD Screen (PC-PTSD; Prins et al., 2003), administered during an initial phone contact with a trained research assistant. Individuals who reported experiencing at least one PTSD Criterion A event and at least one PTSD symptom in the last 30 days were included in this study. The requirement of at least one PTSD symptom was applied to minimize restriction of range in scores on the CAPS-5 and other PTSD measures.

Participants for Sample 1 were recruited into one of three phases: Phase 1 ($n = 31$), Phase 2 ($n = 61$), and Phase 3 ($n = 75$). In Phase 1, CAPS-IV versus CAPS-5 comparisons were based on 30 participants with complete data (original interviewer's ratings) for both interviews; one participant was excluded because he did not complete the CAPS-IV. Interrater reliability analyses were based on 27 participants for CAPS-IV and 28 for CAPS-5 for whom audio recorded interviews were available. For all Phase 1 participants, the index event for symptom inquiry met Criterion A for both *DSM-IV* and *DSM-5* PTSD criteria. In Phase 2, test-retest analyses were based on 60 participants with complete data for both administrations of the CAPS-5; one participant was excluded because he did not complete the CAPS-5.

Participants from Phases 1 through 3 were combined for internal consistency and convergent and discriminant validity analyses. Two participants in Phase 3 did not complete the CAPS-5 and were excluded from all subsequent analyses. Thus, the combined sample for these analyses was 165, including 31 from Phase 1, 61 from Phase 2, and 73 from Phase 3 (see Table 1).

Sample 2 consisted of 207 male veterans who completed the baseline assessment of an ongoing clinical trial (Sloan, Unger, & Gayle Beck, 2016). Eligible veterans were invited to complete an initial assessment (see Sloan et al., 2016 for a detailed overview of study procedures). The only inclusion criteria for the present study were being a male veteran with an index event that met *DSM-5* Criterion A, and self-identifying as being appropriate for a PTSD treatment study. See Table 1 for characteristics of the sample.

Measures: Sample 1. In addition to the CAPS-IV and CAPS-5, the following questionnaire measures were administered to Sample 1 in the order in which they are described in the following sections.

Inventory of Psychosocial Functioning (IPF). The IPF (Marx et al., 2009) is an 80-item self-report measure of functional impairment during the last 30 days in seven specific domains:

Table 1
Descriptive Characteristics of Sample 1 and Sample 2

Characteristic	Sample 1			Sample 2
	Phase 1 ($n = 31$)	Phase 2 ($n = 61$)	Phase 3 ($n = 73$)	($n = 207$)
Age: M , (SD)	52.5 (10.9)	51.2 (12.1)	55.8 (11.7)	55.8 (12.1)
Gender (% Female)	9.7	8.3	18.1	.0
Race (%)				
Caucasian	58.1	65.0	70.4	71.3
Black	32.3	31.7	25.4	19.8
Asian/Pacific Islander	.0	.0	1.4	.0
Native American	.0	.0	.0	2.5
Hispanic/Latino	9.7	3.4	2.8	6.4
Married (%)	16.1	13.6	31.0	43.96
Education: M (SD)	13.7 (2.4)	13.7 (2.1)	14.0 (2.4)	13.9 (2.1)

romantic relationships, family, work, friendships, parenting, education, and self-care. For the first six domains, participants complete items only if the domain applies to them. All participants complete the self-care domain items. Respondents rate the degree to which they have experienced impairment in each area on a 7-point scale from 0 (*never*) to 6 (*always*). Items are summed to total scores for each domain with higher scores indicating greater functional impairment. An overall impairment score is calculated by summing the impairment scores for each domain and then dividing by the number of domains to which the participant responded. IPF scores have demonstrated excellent psychometric properties (Holowka & Marx, 2012). Cronbach's alpha coefficient (α) for IPF domains in the present study ranged from .69 (friendships) to .88 (work).

World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). The WHODAS 2.0 (Ustün et al., 2010) is a 36-item self-report measure of impairment due to health-related problems during the last month across six domains. Respondents rate the degree to which they have experienced impairment on a 5-point scale from *none* to *extreme/cannot do*. Items are summed to total scores for each domain with higher scores indicating greater functional impairment. An overall impairment score is calculated by summing all 36 items. WHODAS 2.0 scores have demonstrated high test-retest reliability and convergent validity (Ustün et al., 2010). In the present study α for global functional disability was .96.

Life Events Checklist (LEC). The LEC (Blake et al., 1990) is a 17-item, self-report measure designed to screen for potentially traumatic events (PTEs) in a respondent's lifetime. The LEC was designed as a companion measure for the CAPS-IV. The LEC assesses exposure to 16 events known to potentially result in PTSD and includes one additional item which allows for a respondent to indicate another extraordinarily stressful event not captured by the first 16 items. For each event, respondents are asked to choose one or more response, including *happened to me*, *witnessed it*, *learned about it*, *not sure*, and *doesn't apply*. The LEC has demonstrated convergent validity with other measures designed to assess exposure to PTEs (Gray, Litz, Hsu, & Lombardo, 2004). In Sample 1, the LEC was used to assess exposure to PTEs and identify the index event; the CAPS was then used to assess whether the index event met Criterion A.

This was a highly traumatized sample: On average, participants endorsed directly experiencing 6.95 PTE categories ($SD = 3.34$).

Using the LEC categories, the most frequently endorsed category was physical assault ($n = 127, 76.0\%$); followed by transportation accident ($n = 120, 71.9\%$); assault with a weapon ($n = 108, 64.7\%$); life threatening illness or injury ($n = 96, 57.5\%$); sudden, unexpected death of someone close ($n = 96, 57.5\%$); natural disaster ($n = 86, 51.5\%$); combat/warzone exposure ($n = 83, 49.7\%$); fire or explosion ($n = 77, 46.1\%$); serious accident ($n = 70, 41.9\%$); exposure to toxic substance ($n = 60, 35.9\%$); unwanted sexual experience other than sexual assault ($n = 58, 34.7\%$); sexual assault ($n = 53, 31.7\%$); causing serious injury, harm, or death to someone else ($n = 44, 26.4\%$); other severe human suffering ($n = 41, 24.6\%$); witnessing sudden, violent death ($n = 33, 19.8\%$); and captivity ($n = 18, 10.8\%$).

PTSD Checklist—Civilian version (PCL-C). The PCL-C (Weathers et al., 1993) is a 17-item, *DSM-IV*-correspondent self-report measure of PTSD. Respondents rate how much they have been bothered by each symptom over the past month using a 5-point scale ranging from 1 (*not at all*) to 5 (*extremely*). For the present study, items were summed to a total score; higher scores indicate greater PTSD symptom severity. PCL-C scores have been extensively validated and have excellent psychometric properties (McDonald & Calhoun, 2010; Wilkins et al., 2011). In the present study alpha was .93.

PTSD Checklist for DSM-5 (PCL-5). The PCL-5 (Weathers, Litz, et al., 2013)—the *DSM-5* revision of the PCL-C—is a 20-item, *DSM-5* correspondent self-report measure of PTSD. Respondents rate how much they have been bothered by each symptom over the past month using a 5-point scale ranging from 0 (*not at all*) to 4 (*extremely*). For the present study, items were summed to a total score; higher scores indicate greater PTSD symptom severity. PCL-5 scores have excellent psychometric properties, including strong internal consistency, test–retest reliability, convergent and discriminant validity, structural validity, diagnostic utility, and sensitivity to clinical change (Blevins, Weathers, Davis, Witte, & Domino, 2015; Bovin et al., 2016; Keane et al., 2014; Wortmann et al., 2016). In the present study alpha was .93.

Psychopathic Personality Inventory—Short Version (PPI-SV). The PPI-SV (Lilienfeld & Andrews, 1996) is a 56-item inventory of the major personality traits of psychopathy in noncriminal populations in eight domains: Machiavellian Egocentricity, Social Potency, Coldheartedness, Carefree Nonplanfulness, Fearlessness, Blame Externalization, Impulsive Nonconformity, and Stress Immunity. Respondents rate the degree to which each statement is true for them on a 4-point scale ranging from 1 (*false*) to 4 (*true*). Items are summed to total scores for each subscale and subscales summed to create a total score; higher scores reflect greater trait psychopathy features. The PPI-SV is based directly on the 187-item Psychopathic Personality Inventory (PPI). PPI scores have demonstrated strong internal consistency, test–retest reliability, and convergent and discriminant validity in several samples of undergraduates (Lilienfeld & Andrews, 1996). In the present study, alpha was .79.

Patient Health Questionnaire (PHQ). The PHQ (Spitzer, Kroenke, & Williams, 1999)—the self-report version of the Primary Care Evaluation of Mental Disorders (PRIME-MD; Spitzer et al., 1994)—is a 58-item self-report measure of current severity of several psychological disorders. Number of response options, response anchors, and time frame vary within and across disorders. Items are summed to create total severity scores for each disorder;

higher scores indicate greater symptom severity. PHQ scores have demonstrated good psychometric properties (Spitzer et al., 1999). In the present study, panic, generalized anxiety, depression, somatoform, and alcohol use disorder scales were used, with alphas of .87, .82, .90, .78, and .72, respectively.

Measures: Sample 2. Along with the CAPS-5, one additional measure, described in the following subsection, was administered to Sample 2 in the present study.

Traumatic Life Events Questionnaire (TLEQ). The TLEQ (Kubany et al., 2000) is a self-report measure designed to assess exposure to 22 potentially traumatic events. Similar to the LEC, the TLEQ also provides one additional item that allows participants to endorse another extremely stressful event not captured by the original 22 items. Each event is scored on a 7-point scale ranging from 0 (*never*) to 6 (*more than five times*); if the event is endorsed, the participant is then asked several follow up questions. The scale has demonstrated good psychometric properties (Kubany et al., 2000). In Sample 2, the TLEQ was used to assess exposure to potentially traumatic events and identify the index event; the CAPS was then used to assess whether the index event met Criterion A. As in Sample 1, trauma exposure was high: The mean number of PTE categories endorsed was 8.90 ($SD = 3.40$). Using the TLEQ categories, the most frequently endorsed category was sudden death of a friend/loved one ($n = 191, 92.3\%$), followed by war zone exposure ($n = 164, 79.2\%$); natural disaster ($n = 149, 72.0\%$); being threatened with serious harm ($n = 119, 57.5\%$); life threatening accident, assault, or illness of a loved one ($n = 109, 52.7\%$); being assaulted by a stranger ($n = 107, 51.7\%$); motor vehicle accident ($n = 101, 48.8\%$); other accident involving severe injury ($n = 100, 48.3\%$); witnessing domestic violence while growing up ($n = 99, 47.8\%$); being robbed ($n = 89, 43.0\%$); experiencing a life-threatening illness ($n = 82, 39.6\%$); childhood physical abuse ($n = 76, 36.7\%$); physically assaulted by intimate partner ($n = 65, 31.4\%$); adult sexual assault ($n = 43, 20.8\%$); being stalked ($n = 41, 19.8\%$); childhood sexual assault (before age 13) by adult ($n = 41, 19.8\%$); childhood sexual assault (before age 13) by similar-aged child ($n = 32, 15.5\%$); and sexual assault between ages 13 and 17 ($n = 18, 8.7\%$).

Procedure.

Sample 1. Participants in Sample 1 were recruited at a VA Healthcare System. Recruited participants either responded to posted flyers or were listed in a large database of veterans who had previously consented to be contacted regarding research participation (following either a clinical evaluation for mental health services or previous research participation) and were contacted by study staff to determine interest in participation in the present study. Institutional review board approval was secured. Eligible participants were consented by study staff. After being consented, participants provided information about their demographics and completed a battery of self-report questionnaires.

After completing the questionnaires, participants were administered either the CAPS-IV or CAPS-5. Participants in Phase 1 and Phase 2 returned for a second visit at which time they were administered either the CAPS-IV or CAPS-5. Participants in Phase 1 were administered the CAPS-IV and the CAPS-5 on separate occasions, in counterbalanced order, by different interviewers blind to all other participant information. Time between assessments ranged from 1 to 6 days ($M = 2.60, SD = 1.13$). Participants in Phase 2 were administered the CAPS-5 at both visits by differ-

ent interviewers blind to all other participant information. Time between assessments again ranged from 1 to 6 days ($M = 2.76$; $SD = 1.09$). Participants in Phase 3 visited the lab only once and were administered the CAPS-5. The CAPS-IV and CAPS-5 were administered by masters- and doctoral-level clinicians with previous training and experience with the CAPS-IV through their clinical and research activities at the National Center for PTSD. Interviewers received CAPS-5 training from the first author and participated in regular calibration meetings which included review of CAPS-5 interviews and discussion of issues regarding standard administration and scoring. Independent ratings of audio-recorded CAPS-IV and CAPS-5 interviews were made by the first and second author, each of whom rated half of the interviews. For all three phases, following participation, participants were compensated monetarily.

Sample 2. Participants in Sample 2 were recruited from VA primary health care and mental health specialty clinics, veterans centers, and flyers posted throughout the VA. Institutional review board approval was secured. Eligible participants provided informed consent and were then administered the CAPS-5 as well as other measures not included in the present study. The CAPS-5 was administered by doctoral-level clinicians who received formal training on the measure. Following participation, participants were compensated monetarily.

Data analysis. Latent variable modeling was conducted using Mplus version 7 (Muthén & Muthén, 1998–2013); all other analyses were conducted using IBM SPSS (Version 22.0). Internal consistency was evaluated with Cronbach's alpha and examination of item-scale total and interitem correlations. These analyses were based on data from the entire Sample 1 (Phases 1 to 3); for Phase 2, the first CAPS-5 administration was used. Interrater reliability and test-retest reliability were evaluated with Cohen's kappa for diagnostic variables and intraclass correlations (ICC) for continuous severity scores, using ICC (1, 1) (Shrout & Fleiss, 1979). Convergent validity of CAPS-5 diagnosis with the CAPS-IV diagnosis was evaluated with kappa. Convergent and discriminant validity were evaluated with Pearson correlations between CAPS-5 total severity score and scores from the questionnaires described earlier. As with internal consistency, these analyses were based on the entire Sample 1, using data from the first CAPS-5 administration in Phase 2. PTSD diagnostic status was determined by applying all DSM diagnostic criteria, including criteria A through F for CAPS-IV and criteria A through G for CAPS-5, and by considering trauma-related ratings for symptoms not inherently linked to the index event.

The latent factor structure of the CAPS-5 was examined using confirmatory factor analysis (CFA), using a combined sample composed of participants in Samples 1 and 2. Because many items did not approximate a normal distribution, items were treated as ordinal (Flora & Curran, 2004; Wirth & Edwards, 2007) and parameters were estimated using the mean- and variance-adjusted weighted least squares (WLSMV) estimator which provides a robust chi-square (Brown, 2006). Review of the covariance matrix indicated a small portion of data was missing (pairwise present data ranged from .96 to 1.00). Missing data were handled using pairwise deletion. Model fit was evaluated using chi-square, Bentler comparative fit index (CFI), Tucker-Lewis Index (TLI), and root mean square error of approximation (RMSEA). Fit statistics were collectively evaluated for each model and established

criteria used to determine good fit ($\chi^2 p \geq .05$, CFI and TLI $\geq .90$, lower limit of the RMSEA 95% confidence interval $< .05$; Bentler, 1990; Brown, 2006; Browne & Cudeck, 1992; Hu & Bentler, 1999; Kline, 2011). Nested models were compared using the DIFFTEST function in Mplus (Muthén & Muthén, 2006), which allows for comparison of nested models using the WLSMV estimator (Brown, 2006).

Results

Internal consistency. Internal consistency was high for CAPS-5 full scale ($\alpha = .88$) and was variable and somewhat lower for the four symptom clusters, including reexperiencing ($\alpha = .77$), avoidance ($\alpha = .55$), NACM ($\alpha = .77$), and alterations in arousal and reactivity ($\alpha = .65$). Because alpha is a function of scale length, the relatively low alpha for avoidance is likely attributable to the fact that this cluster consists of only two items. Mean item-total correlation across all 20 symptoms was .48. Two symptoms, amnesia (D1) and recklessness (E2), had a low item-total correlation (.17 for both). The range of item-total correlations for the remaining 18 symptoms was .37 to .64, with a mean of .51. Most interitem correlations fell in the recommended range of .15 to .50 (Clark & Watson, 1995), with a mean across all 20 symptoms of .26. Amnesia and recklessness also had low interitem correlations, ranging from $-.04$ to .26 for amnesia and $-.03$ to .32 for recklessness. Mean interitem correlation across the remaining 18 symptoms was .29. The low item-total and interitem correlations for amnesia and recklessness are likely attributable to a significant restriction of range owing to very infrequent endorsement of these two symptoms. It may be that these items are important but relatively rare symptoms of PTSD, or it may be that they are simply not representative of the PTSD construct.

Interrater reliability. Interrater reliability (assessed via audio recording) was high. For PTSD diagnosis, kappa was .78 for the CAPS-5, based on the basic SEV2 scoring rule. This was nearly identical to a kappa of .77 for CAPS-IV, based on the basic F1/I2 scoring rule. To calibrate correspondence between CAPS-5 and CAPS-IV (see subsequent text), we added the requirement of a minimum total severity score to SEV2 and F1/I2. We found optimal correspondence (i.e., highest kappa) between CAPS-5 and CAPS-IV using SEV2 plus a total severity score of 26 (SEV2/26) for CAPS-5 and F1/I2 plus a total severity score of 50 (F1/I2/50) for CAPS-IV. Both of these severity scores are in the middle of the moderate range for their respective versions of the CAPS (i.e., 23 to 34 for CAPS-5 and 40 to 59 for CAPS-IV; rationally derived severity score ranges are available from the first author). Interrater reliability was perfect ($\kappa = 1.0$) for both of these scoring rules, with 28 of 28 correct classifications for CAPS-5 SEV2/26 and 27 of 27 correct classifications for CAPS-IV F1/I2/50 (see Table 3 for PTSD prevalence by different CAPS scoring rules).

Finally, interrater reliability for total severity score was high for both CAPS-5 (ICC = .91) and CAPS-IV (ICC = .97). The slightly lower value for CAPS-5 may be due in part to a narrower range of possible severity scores for CAPS-5 (0 to 80) versus for CAPS-IV (0 to 136).

Test-retest reliability. Test-retest reliability for the CAPS-5 (assessed via separate independent interviews) was also high. Kappa was .83 for PTSD diagnosis based on the basic SEV2 scoring rule at Time 1 versus Time 2. This comparison resulted in

Table 2
Descriptive Statistics for CAPS-5 Items

Item	<i>M</i>	<i>SD</i>	Minimum	Maximum	Skewness	Kurtosis	CEP %
B1	2.41	.93	0	4	-1.03	1.16	38.92
B2	1.69	1.33	0	4	-.13	-1.35	33.53
B3	.55	.96	0	3	1.37	.31	21.26
B4	2.12	.98	0	4	-.75	.39	51.80
B5	1.86	1.10	0	4	-.54	-.60	50.00
C1	2.26	1.12	0	4	-.82	-.03	35.33
C2	2.00	1.26	0	4	-.47	-1.00	33.23
D1	.59	1.05	0	4	1.42	.48	17.07
D2	2.02	1.33	0	4	-.46	-1.11	27.84
D3	1.31	1.40	0	4	.42	-1.39	23.35
D4	2.09	1.03	0	4	-.78	-.08	46.41
D5	1.97	1.40	0	4	-.38	-1.30	22.75
D6	2.22	1.26	0	4	-.71	-.66	25.75
D7	1.92	1.38	0	4	-.35	-1.30	25.15
E1	1.41	1.11	0	4	-.06	-1.19	52.10
E2	.41	.89	0	4	1.94	2.42	14.67
E3	2.31	1.18	0	4	-.82	-.30	28.44
E4	1.44	1.14	0	4	-.09	-1.27	49.10
E5	1.88	1.16	0	4	-.45	-.90	43.71
E6	2.36	1.29	0	4	-.77	-.54	21.86

Note. CEP % = clinical elevation prevalence defined as percentage of respondents with item severity scores ≥ 2 .

55 of 60 correct classifications, three with a diagnosis at Time 1 but not Time 2, and two with a diagnosis at Time 2 but not Time 1. Test-retest reliability was slightly lower for SEV2/26 ($\kappa = .73$) but was identical ($\kappa = .83$) for SEV2 plus a total severity score of 23 (SEV2/23), which is the threshold score for the moderate severity score range. Thus, in this sample SEV2 and SEV2 plus the additional requirement of a moderate severity score (SEV2/23) yielded the same high level of test-retest reliability. Test-retest reliability was good for CAPS-5 total severity score (ICC = .78) and adequate-to-good for the four symptom clusters, including reexperiencing (ICC = .80), avoidance (ICC = .67), NACM (ICC = .72), and alterations in arousal and reactivity (ICC = .64).

Convergent and discriminant validity.

CAPS-5 versus CAPS-IV. As expected, we found a moderate association on PTSD diagnosis between the basic CAPS-IV F1/I2 and CAPS-5 SEV2 scoring rules, with a kappa of .51. This comparison resulted in 23 of 30 correct classifications, two with a

diagnosis on the CAPS-IV but not the CAPS-5 and five with a diagnosis on the CAPS-5 but not CAPS-IV. Regarding the two participants with a diagnosis on the CAPS-IV but not the CAPS-5, one was discordant because of not endorsing any symptoms in the NACM cluster on the CAPS-5 (but did endorse both avoidance symptoms on both the CAPS-IV and CAPS-5); the other was discordant because of not endorsing either avoidance symptom on the CAPS-5 (but, responding consistently, did not endorse either avoidance symptom on the CAPS-IV either). Regarding the five participants with a diagnosis on the CAPS-5 but not the CAPS-IV, all were discordant because of endorsing more symptoms in the NACM cluster, by (a) responding inconsistently on corresponding items on the CAPS-IV and CAPS-5, (b) endorsing one or both of the two new items (blame, negative emotions) or the substantially revised negative beliefs item, or (c) both (a) and (b).

Further, as expected, we found a strong association on PTSD diagnosis between CAPS-IV and CAPS-5 when an additional

Table 3
Prevalence of PTSD in Sample 1 and Sample 2

Scoring rule	Sample 1					Sample 2 % (n)
	Phase 1 % (n)	Phase 2 Time 1 % (n)	Phase 2 Time 2 % (n)	Phase 3 % (n)	Full Sample 1	
CAPS-IV F1/I2	33.3 (10)					
CAPS-IV F1/I2/50	30.0 (9)					
CAPS-5 SEV2	43.3 (13)	55.0 (33)	53.3 (32)	58.9 (43)	54.6 (89)	86.5 (179)
CAPS-5 SEV2/23	43.3 (13)	55.0 (33)	53.3 (32)	57.5 (42)	54.0 (88)	83.1 (172)
CAPS-5 SEV2/26	30.0 (9)	53.3 (32)	50.0 (30)	54.8 (40)	49.7 (81)	79.2 (164)

Note. For diagnostic variables, $n = 30$ for Phase 1, $n = 60$ for Phase 2, $n = 73$ for Phase 3, $n = 163$ for full Sample 1, and $n = 207$ for Sample 2. Full Sample 1 prevalence is based on Phase 1, Phase 2 Time 1, and Phase 3. CAPS-IV = Clinician-Administered PTSD Scale for DSM-IV; CAPS-5 = Clinician-Administered PTSD Scale for DSM-5. CAPS-IV F1/I2 = CAPS-IV Frequency = 1/Intensity = 2; CAPS-IV F1/I2/50 = CAPS-IV Frequency = 1/Intensity = 2/Total Severity = 50; CAPS-5 SEV2 = CAPS-5 Item Severity = 2; CAPS-5 SEV2/23 = CAPS-5 Item Severity = 2/Total Severity = 23; CAPS-5 SEV2/26 = CAPS-5 Item Severity = 2/Total Severity = 26.

requirement of a minimum score on total severity score was added to the basic F1/I2 and SEV2 rules. The kappa for CAPS-IV F1/I2/50 versus CAPS-5 SEV2/26 was .84. This comparison resulted in 28 of 30 correct classifications, one with a diagnosis on CAPS-IV but not CAPS-5 and one with a diagnosis on CAPS-5 but not CAPS-IV. This comparison did not result in any new diagnostic discrepancies but succeeded in eliminating five of the seven original discrepancies. The remaining participant with a diagnosis on the CAPS-IV but not the CAPS-5 was the one discordant due to not endorsing any NACM symptoms. The remaining participant with a diagnosis on the CAPS-5 but not the CAPS-IV had a CAPS-5 total severity score of 30 and thus met diagnosis according to the SEV2/26 diagnostic rule. Finally, CAPS-5 total severity score was strongly correlated with CAPS-IV total severity score ($r = .83$).

CAPS-5 and questionnaire measures. Correlations between CAPS-5 scores and the various questionnaire measures were examined to provide evidence of convergent and discriminant validity (see Table 4). CAPS-5 total severity score was most strongly correlated with the PCL-5 and the PCL-C ($r = .66$ for both). Regarding discriminant validity, CAPS-5 total severity score had moderate positive correlations with measures of constructs closely related to PTSD, including anxiety, depression, somatization, disability, and functional impairment ($r_s = .33$ to $.54$). All of these were significantly greater than 0.00 but were significantly lower than .66, that is, the convergent correlation of the CAPS-5 with the PCL-5 and PCL-C, as determined by comparisons of correlated correlations (Meng, Rosenthal, & Rubin, 1992). Finally, CAPS-5 total severity score demonstrated weak, nonsignificant correlations with measures of psychopathy and alcohol abuse ($r_s = .02$ and $.18$, respectively). In general, the CAPS-5 and CAPS-IV demonstrated similar patterns of correlations across the other measures. Differences between them are likely attributable to a substantial difference in sample size: CAPS-5 correlations were based on 165 participants, whereas CAPS-IV correlations were based on only 30 participants, which means the point estimates of CAPS-IV correlations are less reliable and could yield more extreme values.

Latent factor structure. Consistent with Armour et al. (2015), we evaluated six models, including the DSM-5 implicit four-factor model, one additional four-factor model (dysphoria), one five-factor model (dysphoric arousal), two six-factor models (externalizing and anhedonia), and the seven-factor hybrid model. The item mapping for the 20 CAPS-5 items for each of the evaluated models is presented in Table 5, and model fit for each of these models is provided in Table 6. Of the six models examined, each provided generally adequate fit to the data (e.g., the upper limit of the RMSEA 90% CI was below .10 for all six of the models). However, the anhedonia and hybrid models provided the best fit to the data; with the exception of chi-square, each of the fit statistics for these two models met or exceeded established criteria for good fit.

Compared with the DSM-5 model, the dysphoric arousal ($\Delta\chi^2 = 15.80$, $df = 4$, $p < .001$), externalizing ($\Delta\chi^2 = 34.28$, $df = 9$, $p < .001$), anhedonia ($\Delta\chi^2 = 94.90$, $df = 9$, $p < .001$), and hybrid ($\Delta\chi^2 = 120.96$, $df = 15$, $p < .001$) models each provided significantly better fit. The DSM-5 and dysphoria models cannot be compared using chi-square because they are not nested. However, fit statistics were generally comparable for these two models. Compared with the dysphoria model, the dysphoric arousal ($\Delta\chi^2 = 18.89$, $df = 4$, $p < .001$), externalizing ($\Delta\chi^2 = 37.85$, $df = 9$, $p < .001$), anhedonia ($\Delta\chi^2 = 93.36$, $df = 9$, $p < .001$), and hybrid ($\Delta\chi^2 = 120.23$, $df = 15$, $p < .001$) models each provided significantly better fit. Compared with the dysphoric arousal model, the externalizing ($\Delta\chi^2 = 19.65$, $df = 5$, $p < .001$), anhedonia ($\Delta\chi^2 = 74.03$, $df = 5$, $p < .001$), and hybrid ($\Delta\chi^2 = 104.56$, $df = 11$, $p < .001$) models each provided significantly better fit. The externalizing and anhedonia models cannot be compared using chi-square because they are not nested. However, fit statistics were generally stronger for the anhedonia model compared with the externalizing model. Finally, the hybrid model provided significantly better fit compared to both the externalizing ($\Delta\chi^2 = 81.32$, $df = 6$, $p < .001$) and anhedonia ($\Delta\chi^2 = 25.53$, $df = 6$, $p < .001$) models. Collectively, these results indicate that the seven-factor Hybrid model best fit the data.

Table 4
CAPS-5 Convergent and Discriminant Validity Correlations

Measure	1	2	3	4	5	6	7	8	9	10	11
1. CAPS-5	—										
2. CAPS-IV	.83**	—									
3. PCL-5	.66**	.76**	—								
4. PCL-C	.66**	.68**	.91**	—							
5. PHQ-Panic	.33**	.59**	.42**	.45**	—						
6. PHQ-GAD	.47**	.60**	.64**	.65**	.52**	—					
7. PHQ-Depression	.52**	.50**	.68**	.69**	.43**	.74**	—				
8. PHQ-Somatization	.39**	.21	.47**	.52**	.47**	.59**	.58**	—			
9. PHQ-Alcohol Abuse	.18	.26	.13	.10	.26*	.07	.03	-.01	—		
10. PPI	.02	.02	.04	.06	-.05	.03	.05	-.01	.16	—	
11. IPF	.46**	.23	.41**	.48**	.38**	.39**	.53**	.30**	.24*	-.04	—
12. WHODAS 2.0	.54**	.43*	.67**	.69**	.34**	.58**	.75**	.53**	-.03	-.03	.60**

Note. $N = 165$ for CAPS-5 correlations; data includes Sample 1 (Phase 1: $n = 31$, Phase 2: $n = 61$, Phase 3: $n = 73$). CAPS-5 data from Phase 2 is from first administration. $N = 30$ for CAPS-IV correlations; data include Phase 1 of Sample 1, excluding one participant who did not complete the CAPS-IV. CAPS-5 = Clinician Administered PTSD Scale for DSM-5; CAPS-IV = Clinician Administered PTSD Scale for DSM-IV; PCL-5 = PTSD Checklist for DSM-5; PCL-C = PTSD Checklist for DSM-IV—Civilian Version; PHQ = Patient Health Questionnaire; GAD = generalized anxiety disorder; PPI = Psychopathic Personality Inventory; IPF = Inventory of Psychosocial Functioning; WHODAS 2.0 = World Health Organization Disability Assessment Schedule Version 2.0.

* $p < .05$. ** $p < .01$.

Table 5
Item Mapping for CAPS-5 Measurement Models

CAPS-5 Item	Item description	Model					
		DSM-5	Dysphoric arousal	Dysphoria	Externalizing behaviors	Anhedonia	Hybrid
1 (B1)	Memories	INT	INT	INT	INT	INT	INT
2 (B2)	Dreams	INT	INT	INT	INT	INT	INT
3 (B3)	Flashbacks	INT	INT	INT	INT	INT	INT
4 (B4)	Cued distress	INT	INT	INT	INT	INT	INT
5 (B5)	Cued physical reactions	INT	INT	INT	INT	INT	INT
6 (C1)	Avoiding internal reminders	AVD	AVD	AVD	AVD	AVD	AVD
7 (C2)	Avoiding external reminders	AVD	AVD	AVD	AVD	AVD	AVD
8 (D1)	Dissociative amnesia	NCM	NCM	DYS	NCM	NAF	NAF
9 (D2)	Negative beliefs	NCM	NCM	DYS	NCM	NAF	NAF
10 (D3)	Blame	NCM	NCM	DYS	NCM	NAF	NAF
11 (D4)	Negative feelings	NCM	NCM	DYS	NCM	NAF	NAF
12 (D5)	Loss of interest	NCM	NCM	DYS	NCM	ANH	ANH
13 (D6)	Detachment or estrangement	NCM	NCM	DYS	NCM	ANH	ANH
14 (D7)	Numbing	NCM	NCM	DYS	NCM	ANH	ANH
15 (E1)	Irritability or aggressive behavior	AAR	DAR	DYS	EXT	DAR	EXT
16 (E2)	Reckless behavior	AAR	DAR	DYS	EXT	DAR	EXT
17 (E3)	Hypervigilance	AAR	AXA	AXA	AXA	AXA	AXA
18 (E4)	Startle	AAR	AXA	AXA	AXA	AXA	AXA
19 (E5)	Concentration	AAR	DAR	DYS	DAR	DAR	DAR
20 (E6)	Sleep	AAR	DAR	DYS	DAR	DAR	DAR

Note. CAPS-5 = Clinician-Administered PTSD Scale for DSM-5; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders (5th ed.); INT = Intrusions cluster; AVD = Avoidance cluster; NCM = Negative alterations in cognition and mood cluster; AAR = Alterations in arousal and reactivity cluster; DAR = dysphoric arousal cluster; AXA = anxious arousal cluster; DYS = dysphoria cluster; EXT = Externalizing cluster; ANH = Anhedonia cluster.

Within the hybrid model, all items had significant loadings onto their respective latent variables. The magnitude of these loadings was salient (i.e., standardized parameter estimates of .3 or greater; Brown, 2006) for all items. Item D1 (dissociative amnesia; see Table 7) had a relatively low-magnitude loading. This finding suggests this item is not a strong indicator of the negative affect latent variable. Of note, this symptom was also among those with the lowest clinical elevation prevalence (see Table 2). Accordingly, restriction of range may have contributed to this relatively weak loading. Aside from this item, loadings for other items indicated that all other symptoms are good indicators of their respective symptom clusters.

Discussion

In this article, we presented the results of a comprehensive psychometric evaluation of CAPS-5 scores in two samples of

military veterans. All hypotheses were supported. First, CAPS-5 scores demonstrated strong internal consistency, interrater reliability, and test-retest reliability. This indicates that CAPS-5 scores reflect relatively little measurement error due to items, raters, or occasions. Second, CAPS-5 total severity score was strongly correlated with the CAPS-IV, PCL-5, and PCL-C moderately correlated with measures of depression, anxiety, somatization, and functional impairment; and weakly and nonsignificantly correlated with measures of alcohol abuse and psychopathy. This indicates good construct validity in that the CAPS-5 scores demonstrated a conceptually consistent pattern of associations with a wide range of external variables.

Third, CAPS-5 PTSD diagnosis was moderately associated with CAPS-IV diagnosis when the basic CAPS-5 SEV2 and CAPS-IV F1/I2 scoring rules were used, but strongly associated when the requirement of a minimum total severity score was added (26 for

Table 6
Fit Statistics of CAPS-5 Measurement Models

Model	χ^2	df	p	CFI	TLI	RMSEA (90% CI)
DSM-5	394.19	164	<.001	.95	.94	.07 (.06-.07)
Dysphoria	397.63	164	<.001	.94	.94	.07 (.06-.07)
Dysphoric arousal	381.72	160	<.001	.95	.94	.07 (.06-.07)
Externalizing	366.01	155	<.001	.95	.94	.06 (.06-.07)
Anhedonia	291.65	155	<.001	.97	.96	.05 (.04-.06)
Hybrid	267.85	149	<.001	.97	.96	.05 (.04-.06)

Note. CAPS-5 = Clinician-Administered PTSD Scale for DSM-5; CFI = Bentler Comparative Fit Index; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, 5th Edition; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation.

Table 7
Standardized and Unstandardized Parameter Estimates for the CAPS-5 Seven-Factor Hybrid Measurement Model

PTSD symptom	Factor	Estimate	SE	STDYX
1. Intrusive memories	Intrusions	1.00*	.00	.82
2. Nightmares		.60*	.06	.49
3. Flashbacks		.58*	.08	.48
4. Cued distress		.95*	.05	.78
5. Cued physical reactions		1.03*	.05	.85
6. Avoidance of thoughts	Avoidance	1.00*	.00	.71
7. Avoidance of reminders		1.03*	.07	.73
8. Trauma-related amnesia	Negative affect	1.00*	.00	.34
9. Negative beliefs		2.02*	.41	.69
10. Blame		1.58*	.33	.53
11. Negative feelings		2.24*	.46	.76
12. Loss of interest	Anhedonia	1.00*	.00	.78
13. Feeling detached		1.11*	.05	.86
14. Feeling numb		1.05*	.05	.82
15. Irritability	Externalizing	1.00*	.00	.68
16. Risk taking		.70*	.16	.48
17. Hypervigilance	Anxious arousal	1.00*	.00	.75
18. Startle		.74*	.08	.56
19. Difficulty concentrating	Dysphoric arousal	1.00*	.00	.55
20. Sleep disturbance		.98*	.12	.54

Note. CAPS-5 = Clinician-Administered PTSD Scale for *DSM-5*; Estimate = unstandardized parameter estimates; SE = standard error of the unstandardized parameter estimates; STDYX = Standardized parameter estimates.

* $p < .05$.

SEV2 score and 50 for F1/I2). This indicates strong backward compatibility between CAPS-5 and CAPS-IV when they are optimally calibrated. Fourth, interrater reliability was good for both SEV2 and F1/I2 and perfect for SEV2/26 and F1/I2/50, indicating little measurement error due to rater. Last, CFA indicated that the four-factor *DSM-5* model provided adequate fit to the CAPS-5 data, but that the seven-factor Hybrid model provided good fit and was the best-fitting model overall. Thus, the present study replicated previous self-report CFA studies of the *DSM-5* PTSD criteria and extended findings to the structured interview format.

DSM-5 versus DSM-IV criteria.

Correspondence of CAPS-5 with CAPS-IV. A key issue we addressed is the comparability of *DSM-5* and *DSM-IV* PTSD diagnostic criteria. Our findings indicate that CAPS-5 and CAPS-IV diagnoses closely correspond when properly calibrated. Accordingly, the great majority of participants are classified the same despite the *DSM-5* revisions to both the PTSD criteria and the CAPS. Further, the level of discordance was nearly identical to that found for test-retest reliability of the CAPS-5. This indicates that the discordance between *DSM-5* and *DSM-IV* is no greater than the discordance found between independent administrations of the CAPS-5, which could result either from measurement error (due to occasions, interviewers, or the interaction of occasions and interviewer) or actual changes in participants' diagnostic status in the retest interval.

Sources of diagnostic discordance. The new *DSM-5* requirement of at least one avoidance symptom is one potential source of diagnostic discrepancy because it is possible to meet the *DSM-IV* C criterion without any avoidance symptoms. In the National Stressful Events Survey Kilpatrick et al. (2013) identified the new avoidance requirement (along with exclusion of nonviolent death

as a Criterion A event) as one of the two main sources of discrepancy between *DSM-IV* and *DSM-5*. We did not replicate Kilpatrick et al.'s (2013) finding. Instead, in the initial CAPS-5 SEV2 versus CAPS-IV F1/I2 comparison we found that only one participant had a diagnostic discrepancy because of failure to meet the new avoidance requirement, and even this one participant was no longer discrepant in the calibrated SEV2/26 versus F1/I2/50 comparison.

There are several methodological differences between the present study and the Kilpatrick et al. study that might account for this failure to replicate, including sample size, population, and instrumentation. For example, the Kilpatrick et al. study involved a very large ($N = 2,953$) national sample of adults, whereas the present study involved a small convenience sample of veterans, so the failure to replicate may be due to inadequate power or differences in populations. Perhaps more importantly, though, the Kilpatrick et al. study used a self-report survey instrument, whereas we used clinical interviews, so the failure to replicate may be due to assessment modality. Given that the source of the discrepancy involves the avoidance symptoms, which are negative (deficit) symptoms, it is plausible that respondents may be less aware of their avoidance and thus less likely to report it on self-report measures—and conversely more likely to report them on interviews when prompted by the interviewer. More research is needed to determine whether the new avoidance requirement is a significant source of diagnostic discrepancy between *DSM-5* and *DSM-IV* PTSD and under what conditions it might be observed. In any case, in the present study the new avoidance requirement was not a source of diagnostic discrepancy.

Diagnostic discrepancies between CAPS-5 and CAPS-IV in the initial SEV2 versus F1/I2 comparison were primarily due to par-

participants endorsing more items on the CAPS-5, specifically from the *DSM-5* NACM cluster. This resulted either from participants giving different responses to symptoms that appear on both the CAPS-5 and CAPS-IV (e.g., detachment/estrangement) or from participants endorsing additional symptoms that only appear on the CAPS-5 (i.e., the new *DSM-5* blame and negative emotions symptoms or the negative beliefs symptom, a substantially expanded version of the *DSM-IV* foreshortened future symptom). Instances of participants giving different responses to NACM symptoms that appear on both the CAPS-5 and CAPS-IV occurred despite prompts and scoring for these items that are nearly identical on the CAPS-IV and CAPS-5. This pattern suggests that participants were responding inconsistently across occasions to essentially the same prompts, a common source of measurement error in test-retest reliability studies.

CAPS-5 versus PSSI-5. As noted in the introduction, Foa et al. (2016) found only moderate correspondence between PSSI-5 and CAPS-5 diagnostic status, with the PSSI-5 having a sensitivity of .82, specificity of .71, and kappa of .49 against the CAPS-5 as criterion. The relatively low specificity suggests that the standard PSSI-5 scoring rule—whereby a symptom is considered present when an item is rated as 1 = *once per week or less/a little*—is more lenient than the CAPS-5 SEV2 rule, yielding a 29% false positive rate. However, the standard PSSI-5 scoring rule also yielded an 18% false negative rate, suggesting that simply applying a more stringent PSSI-5 rule would not necessarily improve correspondence with the CAPS-5 because it might increase false negatives.

There are several possible reasons there was only moderate correspondence between the PSSI-5 and CAPS-5. First, this might be a study-specific finding due to a relatively small sample size or some idiosyncratic aspects of the specific participants or interviewers involved. Better correspondence might be obtained with a larger sample in a different context. Second, given substantial differences in how symptom information is quantified on the PSSI-5 versus the CAPS-5, it might be that the two interviews just need to be optimally calibrated, much as we did with the CAPS-IV and CAPS-5. Foa et al. (2016) conducted one type of calibration, using an ROC analysis to identify the optimal PSSI-5 total severity score for predicting a CAPS-5 diagnosis. However, this approach still yielded only moderate correspondence, with a sensitivity of .77 and specificity of .77. Third, correspondence between the PSSI-5 and CAPS-5 may be reduced because of less than perfect reliability in one or both measures, particularly with respect to test-retest reliability, which indicates the reproducibility or stability of diagnostic status across testing occasions. Foa et al. did not report reliability for CAPS-5 data. However, they reported a test-retest kappa of .65 for the PSSI-5, which, although in the good range, indicates at least a moderate amount of diagnostic unreliability, which could attenuate correspondence of the PSSI-5 with external correlates such as the CAPS-5. Additional head-to-head comparisons of the PSSI-5 and CAPS-5 are needed to investigate these various possibilities.

Limitations and Conclusion

Our study has several important limitations. First, all participants were military veterans recruited from a single geographical region, and the great majority were men. Accordingly, it is unclear

how well the results generalize to nonveterans and women. Second, the sample sizes were modest, especially for the comparison between CAPS-5 and CAPS-IV in Phase 1. Our aim for this phase was simply to demonstrate the backward compatibility of the CAPS-5 with the CAPS-IV to provide context for a more intensive focus on the reliability and validity of the CAPS-5. When we designed the study and implemented data collection, the relatively small sample size seemed adequate for the task, as it proved to be. However, a larger sample would have provided an even more convincing demonstration. Third, to keep the study protocol manageable, we used a relatively limited number of external correlates to examine convergent and discriminant validity. Clearly, more studies need to be conducted using the CAPS-5 and a much wider range of validity evidence, including other interviews, self-report measures, behavioral observations, physiological measures, and response to treatment.

Despite these limitations, our study provides clear evidence that the CAPS-5 is a psychometrically sound measure of *DSM-5* PTSD diagnostic status and symptom severity. In addition, our experience is that the streamlined format for CAPS-5 facilitates administration and scoring and makes it easier to learn than CAPS-IV. Finally, the carryover features, including carefully worded prompts, behaviorally anchored ratings, and trauma-related inquiry for individual symptoms, ensure that the CAPS-5 retains the best aspects of previous versions of the CAPS. Thus, the CAPS-5 is linked to the extensive validation literature regarding the CAPS-IV and thereby provides continuity in evidence-based assessment of PTSD as the field of traumatic stress transitions from *DSM-IV* to *DSM-5* criteria.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Armour, C., Tsai, J., Durham, T. A., Charak, R., Biehn, T. L., Elhai, J. D., & Pietrzak, R. H. (2015). Dimensional structure of *DSM-5* posttraumatic stress symptoms: Support for a hybrid Anhedonia and Externalizing Behaviors model. *Journal of Psychiatric Research*, *61*, 106–113. <http://dx.doi.org/10.1016/j.jpsychires.2014.10.012>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246. <http://dx.doi.org/10.1037/0033-2909.107.2.238>
- Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Klauminzer, G., Charney, D. S., & Keane, T. M. (1990). A clinician rating scale for assessing current and lifetime PTSD: The CAPS-1. *Behavior Therapist*, *13*, 187–188.
- Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, *28*, 489–498. <http://dx.doi.org/10.1002/jts.22059>
- Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. *Psychological Assessment*, *28*, 1379–1391. <http://dx.doi.org/10.1037/pas0000254>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*, 230–258. <http://dx.doi.org/10.1177/0049124192021002005>

- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319. <http://dx.doi.org/10.1037/1040-3590.7.3.309>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart, & Winston.
- Elhai, J. D., Gray, M. J., Kashdan, T. B., & Franklin, C. L. (2005). Which instruments are most commonly used to assess traumatic event exposure and posttraumatic effects? A survey of traumatic stress professionals. *Journal of Traumatic Stress, 18*, 541–545. <http://dx.doi.org/10.1002/jts.20062>
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491. <http://dx.doi.org/10.1037/1082-989X.9.4.466>
- Foa, E. B., McLean, C. P., Zang, Y., Zhong, J., Rauch, S., Porter, K., . . . Kauffman, B. Y. (2016). Psychometric properties of the Posttraumatic Stress Disorder Symptom Scale Interview for *DSM-5* (PSSI-5). *Psychological Assessment, 28*, 1159–1165. <http://dx.doi.org/10.1037/pas0000259>
- Friedman, M. J. (2013). Finalizing PTSD in *DSM-5*: Getting here from there and where to go next. *Journal of Traumatic Stress, 26*, 548–556. <http://dx.doi.org/10.1002/jts.21840>
- Gray, M., Litz, B., Hsu, J., & Lombardo, T. (2004). Psychometric properties of the Life Events Checklist. *Assessment, 11*, 330–341. <http://dx.doi.org/10.1177/1073191104269954PILOTSID:26825>
- Haynes, S. N., Richard, D. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238–247. <http://dx.doi.org/10.1037/1040-3590.7.3.238>
- Holowka, D. W., & Marx, B. P. (2012). Assessing PTSD-related functional impairment and quality of life. In J. Beck & D. M. Sloan (Eds.), *The Oxford handbook of traumatic stress disorders* (pp. 315–330). New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780195399066.013.0021>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- Keane, T. M., Caddell, J. M., & Taylor, K. A. (1988). The Mississippi Scale for Combat Related PTSD: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology, 56*, 85–90. <http://dx.doi.org/10.1037/0022-006X.56.1.85>
- Keane, T. M., Rubin, A., Lachowicz, M., Brief, D., Enggasser, J. L., Roy, M., . . . Rosenbloom, D. (2014). Temporal stability of *DSM-5* posttraumatic stress disorder criteria in a problem-drinking sample. *Psychological Assessment, 26*, 1138–1145. <http://dx.doi.org/10.1037/a0037133>
- Kilpatrick, D. G., Resnick, H. S., Milanak, M. E., Miller, M. W., Keyes, K. M., & Friedman, M. J. (2013). National estimates of exposure to traumatic events and PTSD prevalence using *DSM-IV* and *DSM-5* criteria. *Journal of Traumatic Stress, 26*, 537–547. <http://dx.doi.org/10.1002/jts.21848>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Kubany, E. S., Haynes, S. N., Leisen, M. B., Owens, J. A., Kaplan, A. S., Watson, S. B., & Burns, K. (2000). Development and preliminary validation of a brief broad-spectrum measure of trauma exposure: The Traumatic Life Events Questionnaire. *Psychological Assessment, 12*, 210–224. <http://dx.doi.org/10.1037/1040-3590.12.2.210>
- Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment, 66*, 488–524. http://dx.doi.org/10.1207/s15327752jpa6603_3
- Marmar, C. R., Schlenger, W., Henn-Haase, C., Qian, M., Purchia, E., Li, M., . . . Kulka, R. A. (2015). Course of posttraumatic stress disorder 40 years after the Vietnam War: Findings from the National Vietnam Veterans Longitudinal Study. *Journal of the American Medical Association Psychiatry, 72*, 875–881. <http://dx.doi.org/10.1001/jamapsychiatry.2015.0803>
- Marx, B. P., Schnurr, P. P., Rodriguez, P., Holowka, D. W., Lunney, C., Weathers, F., . . . Keane, T. M. (2009, November). *Development and validation of a scale to assess functional impairment among active duty service members and veterans*. Paper presented at the 25th Annual Meeting of the International Society for Traumatic Stress Studies, Atlanta, GA.
- McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD checklist: A critical review. *Clinical Psychology Review, 30*, 976–987. <http://dx.doi.org/10.1016/j.cpr.2010.06.012>
- Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*, 172–175. <http://dx.doi.org/10.1037/0033-2909.111.1.172>
- Muthen, B. O., & Muthen, L. K. (2006). *Chi-square difference testing using the Satorra-Bentler scaled chi-square*. Retrieved from <http://statmodel.com/chidiff.shtml>
- Muthén, L. K., & Muthén, B. O. (1998–2013). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.
- Prins, A., Ouimette, P., Kimerling, R., Cameron, R. P., Hugelshofer, D. S., Shaw-Hegwer, J., . . . Sheikh, J. I. (2003). The Primary Care PTSD Screen (PC-PTSD): Development and operating characteristics. *Primary Care Psychiatry, 9*, 9–14. <http://dx.doi.org/10.1185/135525703125002360>
- Schnurr, P. P., Spiro, A., III, Vielhauer, M. J., Findler, M. N., & Hamblen, J. L. (2002). Trauma in the lives of older men: Findings from the Normative Aging Study. *Journal of Clinical Geropsychology, 8*, 175–187. <http://dx.doi.org/10.1023/A:1015992110544>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428. <http://dx.doi.org/10.1037/0033-2909.86.2.420>
- Sloan, D. M., Unger, W., & Gayle Beck, J. (2016). Cognitive-behavioral group treatment for veterans diagnosed with PTSD: Design of a hybrid efficacy-effectiveness clinical trial. *Contemporary Clinical Trials, 47*, 123–130. <http://dx.doi.org/10.1016/j.cct.2015.12.016>
- Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. Primary Care Evaluation of Mental Disorders. *Journal of the American Medical Association, 282*, 1737–1744. <http://dx.doi.org/10.1001/jama.282.18.1737>
- Spitzer, R. L., Williams, J. B., Kroenke, K., Linzer, M., deGruy, F. V., III, Hahn, S. R., . . . Johnson, J. G. (1994). Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *Journal of the American Medical Association, 272*, 1749–1756. <http://dx.doi.org/10.1001/jama.1994.03520220043029>
- Ustün, T. B., Chatterji, S., Kostanjsek, N., Rehm, J., Kennedy, C., Epping-Jordan, J., . . . the WHO/NIH Joint Project. (2010). Developing the World Health Organization Disability Assessment Schedule 2.0. *Bulletin of the World Health Organization, 88*, 815–823. <http://dx.doi.org/10.2471/BLT.09.067231>
- Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013). *The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)*. Retrieved from www.ptsd.va.gov
- Weathers, F. W., Keane, T. M., & Davidson, J. R. T. (2001). Clinician-administered PTSD scale: A review of the first ten years of research. *Depression and Anxiety, 13*, 132–156. <http://dx.doi.org/10.1002/da.1029>
- Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD Checklist: Reliability, validity, and diagnostic utility*. Paper presented at the Annual Meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

- Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD Checklist for DSM-5 (PCL-5)*. Retrieved from www.ptsd.va.gov
- Weathers, F. W., Marx, B. P., Friedman, M. J., & Schnurr, P. P. (2014). Posttraumatic stress disorder in *DSM-5*: New criteria, new measures, and implications for assessment. *Psychological Injury and Law, 7*, 93–107. <http://dx.doi.org/10.1007/s12207-014-9191-1>
- 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. <http://dx.doi.org/10.1037/1040-3590.11.2.124>
- Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . . the QUADAS-2 Group. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*, 529–536. <http://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009>
- Wilkins, K. C., Lang, A. J., & Norman, S. B. (2011). Synthesis of the psychometric properties of the PTSD checklist (PCL) military, civilian, and specific versions. *Depression and Anxiety, 28*, 596–606. <http://dx.doi.org/10.1002/da.20837>
- Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*, 58–79. <http://dx.doi.org/10.1037/1082-989X.12.1.58>
- Wortmann, J. H., Jordan, A. H., Weathers, F. W., Resick, P. A., Dondanville, K. A., Hall-Clark, B., . . . Litz, B. T. (2016). Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. *Psychological Assessment, 28*, 1392–1403. <http://dx.doi.org/10.1037/pas0000260>

Received August 3, 2016

Revision received March 16, 2017

Accepted March 16, 2017 ■