Development and Evaluation of the Dutch Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)

Manon A. Boeschoten, Niels Van der Aa, Anne Bakker, F. Jackie June Ter Heide, Marthe C. Hoofwijk, Ruud A. Jongedijk, Agnes Van Minnen, Bernet M. Elzinga & Miranda Olff

EUROPEAN JOURNAL OF PSYCHO-TRAUMATOLOGY
THE OFFICIAL JOURNAL OF THE EUROPEAN SOCIETY FOR TRAUMATIC STRESS STUDIES

Taylor & Francis
Taylor & Francis Group

CLINICAL RESEARCH ARTICLE

Development and Evaluation of the Dutch Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)

Manon A. Boeschoten[a], Niels Van der Aa[a], Anne Bakker [b], F. Jackie June Ter Heide[a], Marthe C. Hoofwijk[a,c], Ruud A. Jongedijk[a,d], Agnes Van Minnen[e], Bernet M. Elzinga[f] and Miranda Olff[b,d]

[a]Foundation Centrum '45, Oegstgeest/Diemen, The Netherlands | Partner in Arq Psychotrauma Expert Group, Diemen, The Netherlands; [b]Department of Psychiatry, Academic Medical Center, Amsterdam, The Netherlands; [c]Psychotrauma Diagnosis Center, Diemen, The Netherlands | partner in Arq Psychotrauma Expert Group, Diemen, The Netherlands; [d]Arq Psychotrauma Expert Group, Diemen, The Netherlands; [e]PSYTREC, The Netherlands & Radboud University, Nijmegen, The Netherlands; [f]Department of Clinical Psychology, Leiden University, Leiden, The Netherlands & Leiden Institute for Brain and Cognition, Leiden, The Netherlands

ABSTRACT

Background: In 2013, the Clinician-Administered PTSD Scale, the golden standard to assess PTSD, was adapted to the DSM-5 (CAPS-5).

Objective: This project aimed to develop a clinically relevant Dutch translation of the CAPS-5 and to investigate its psychometric properties.

Method: We conducted a stepped translation including Delphi rounds with a crowd of 44 Dutch psychotrauma experts and five senior psychotrauma experts. Using partial crowd-translations, two professional translations and the official Dutch translation of the DSM-5, each senior expert aggregated one independent translation. Consensus was reached plenary. After back-translation, comparison with the original CAPS-5 and field testing, a last round with the senior experts resulted in the final version. After implementation clinicians conducted CAPS-5 interviews with 669 trauma-exposed individuals referred for specialized diagnostic assessment. Reliability of the Dutch CAPS-5 was investigated through internal consistency and interrater reliability analyses, and construct validity through confirmatory factor analysis (CFA).

Results: CAPS-5 total severity score showed high internal consistency ($\alpha = .90$) and interrater reliability (ICC = .98, 95% CI: .94–.99). CAPS-5 diagnosis showed modest interrater reliability (kappa = .59, 95% CI: .20–.98). CFA with alternative PTSD models revealed adequate support for the DSM-5 four-factor model, but a six-factor (Anhedonia) model fit the data best.

Conclusions: The Dutch CAPS-5 is a carefully translated instrument with adequate psychometric properties. Current results add to the growing support for more refined (six and seven) factor models for DSM-5 PTSD indicating that the validity and clinical implications of these models should be objective of further research.

HIGHLIGHTS
• Novel stepped crowd-translation and psychometric evaluation of golden standard clinical interview for PTSD; CAPS-5.
• Carefully translated Dutch CAPS-5 with adequate measurement properties.
• Further evidence for more refined factor models for DSM-5 PTSD.

Desarrollo y Evaluación de la Escala Holandesa para el TEPT Aplicada por el Clínico adaptada al DSM-5 (CAPS-5)
RESUMEN

Antecedentes: En el año 2013 la Escala para el TEPT Aplicada por el Clínico, la prueba estándar para evaluar el TEPT, fue adaptada al DSM-5 (CAPS-5).

Objetivo: Este proyecto apuntó a desarrollar una traducción holandesa clínicamente relevante de la Escala para el TEPT Aplicada por el Clínico adaptada al DSM-5 (CAPS-5) e investigar sus propiedades psicométricas.

Método: Realizamos una traducción escalonada, incluyendo fases del método Delphi con un grupo de 44 expertos holandeses en psicotrauma y cinco expertos de larga trayectoria en psicotrauma. Utilizando traducciones en grupo parciales, dos traducciones profesionales y la traducción holandesa oficial del DSM-5, cada experto experimentado sumó una traducción independiente. Se alcanzó un conceso pleno. Después de traducciones inversas, comparación con el CAPS-5 original y ensayo de campo, una última fase con los expertos experimentados resultó en la versión definitiva. Tras la implementación, los clínicos realizaron entrevistas aplicando CAPS-5 a 669 individuos expuestos a trauma referidos por evaluación diagnóstica especializada. Se investigó la fiabilidad del CAPS-5 holandés a través de consistencia interna y análisis de confiabilidad, y se estableció su validez a través de análisis factorial de tipo confirmatorio.

Resultados: El puntaje de severidad total del CAPS-5 mostró alta consistencia interna ($\alpha = .90$) y confiabilidad (ICC = .98, 95% IC: .94 - .99). El diagnóstico de CAPS-5 mostró una modesta confiabilidad (kappa = .59, 95% CI: .20 - .98). El análisis factorial de tipo confirmatorio con

modelos alternativos de TEPT reveló un respaldo adecuado para el modelo de 4 factores del DSM-5, pero un modelo de 6 factores (Anhedonia) se ajusta mejor a los datos.
Conclusiones: El CAPS-5 holandés es un instrumento cuidadosamente traducido con adecuadas propiedades psicométricas. Nuestros resultados se suman al respaldo creciente para modelos de factores (seis y siete) más refinados para el TEPT según el DSM-5, indicando que la validez y las implicaciones clínicas de estos modelos deberían ser objeto de futuras investigaciones.

荷兰语版DSM-5 PTSD临床使用量表（CAPS-5）的开发和评估

背景：2013年，DSM-5 PTSD临床使用量表（评估PTSD的黄金标准）经过修订适用于DSM-5（CAPS-5）。
目标：该项目旨在开发荷兰语版本的临床CAPS-5，并考察其心理测量学特性。
方法：我们采用了一个分步的翻译方法，包括44名荷兰精神创伤专家和5名资深精神创伤专家的Delphi rounds流程。使用不完全'人群翻译法'（crowd-translations），由两个专业翻译和DSM-5官方的荷兰语翻译进行翻译，每一位资深专家将汇总出一个独立的翻译版本。之后通过全体会议确定一个共同版本。经过反向翻译、与原CAPS-5的比较和田野测试后，在最后一轮中资深专家确定了最终版本。临床医生对669名创伤暴露后需要进行专门诊断评估的个体进行了CAPS-5访谈。通过内部一致性和评分者一致性分析考察荷兰语版CAPS-5的信度，并通过验证性因子分析（CFA）考察其效度。
结果：CAPS-5严重度总分显示高度的内部一致性（α= .90）和评分者一致性（ICC = .98,95％CI：.94- .99）。CAPS-5诊断显示中等的评分者一致性（kappa = .59,95％CI：.20 - .98）。对PTSD替代模型进行CFA结果充分支持DSM-5四因子模型，但六因子（Anhedonia）模型最拟合数据。
结论：荷兰语版CAPS-5是一个经过精心翻译的工具，具有足够良好的心理测量特性。本研究结果支持了DSM-5 PTSD更精细的（6和7）因子模型，表明这些模型的有效性和临床意义影视是进一步研究的目标。

1. PTSD in DSM-5

Posttraumatic Stress Disorder (PTSD) in the *Diagnostic and Statistical Manual for Mental Disorders 5th edition* (DSM-5; American Psychiatric Association [APA], 2013) is characterized by symptoms of intrusion, avoidance, negative alterations in cognitions and mood, and alterations in arousal and reactivity, following exposure to a traumatic event, along with significant functional impairment and at least a one-month duration of symptoms. The DSM-5 has adopted some critical changes to the diagnostic criteria for PTSD in comparison to DSM-IV (APA, 1994) both on an item and a cluster level (see Friedman, 2013), based on scientific insights regarding the symptomatology and factor structure of PTSD. The stressor criterion (A) was refined and the requirement of experiencing fear, helplessness, or horror during the event was eliminated. The DSM-IV avoidance and numbing cluster was divided into two separate clusters: Avoidance (C) and Alterations in cognitions and mood (D). Three new symptoms were introduced: Distorted cognitions leading to blaming self or others (D3), Persistent negative emotional state (D4), and Reckless and self-destructive behaviour (E2), reflecting the insight that PTSD symptomatology goes beyond merely anxiety symptoms. The old Sense of a foreshortened future item has been reformulated into Persistent and exaggerated negative beliefs or expectations about oneself, others, or the world (D2). In DSM-5, it is explicitly required that all symptoms have begun or worsened after the traumatic event.

Following the changes in diagnostic criteria for PTSD in DSM-5, clinical instruments to assess PTSD have also been updated. The golden standard to assess PTSD is the *Clinician-Administered PTSD Scale* (CAPS), which was adapted to the DSM-5 by Weathers and colleagues (Olff, 2015; Weathers et al., 2013a). In addition, changes were introduced in the layout and scoring of the CAPS-5 in order to streamline administration and simplify scoring (Weathers et al., 2017). Psychometric evaluation of the CAPS-5 demonstrated high internal consistency, strong interrater reliability and test–retest reliability, and strong correspondence with a diagnosis based on the CAPS for DSM-IV (Weathers et al., 2017). Convergent and discriminant validity was confirmed by a strong correlation with the PTSD Checklist for DSM-5 (PCL-5; Blevins, Weathers, Davis, Witte, & Domino, 2015; Weathers et al., 2013b) and other measures of psychopathology and functioning (Weathers et al., 2017). To enable comparisons between international trauma studies and to further the field of psychotrauma on a global level, proper translations of revised PTSD instruments such as the CAPS-5 are required and their psychometric properties need to be evaluated (Weathers et al., 2017; Weathers, Marx, Friedman, & Schnurr, 2014).

2. Factor structure of PTSD

The latent factor structure of PTSD has been extensively debated for over two decades and various models have been proposed and tested (Armour, Ross, & Elhai, 2016a; Elhai & Palmieri, 2011b; Rademaker et al., 2012). For an illustration of the most common PTSD models and their symptom mapping across factors see Table 1 (adapted from Armour et al., 2016a). Before the release of DSM-5,

Table 1. Symptom mappings of the most common DSM-5 PTSD models.

Symptom	DSM-5 Numbing four-factor	Dysphoria four-factor	Dysphoric arousal five-factor	Externalizing behaviours six-factor	Anhedonia six-factor	Hybrid seven-factor
B1: Intrusive thoughts	R	R	R	R	R	R
B2: Nightmares	R	R	R	R	R	R
B3: Flashbacks	R	R	R	R	R	R
B4: Psychological cue reactivity	R	R	R	R	R	R
B5: Physiological cue reactivity	R	R	R	R	R	R
C1: Avoidance of thoughts	AV	AV	AV	AV	AV	AV
C2: Avoidance of reminders	AV	AV	AV	AV	AV	AV
D1: Memory impairment	NACM	D	NACM	NACM	NA	NA
D2: Negative beliefs	NACM	D	NACM	NACM	NA	NA
D3: Distorted blame	NACM	D	NACM	NACM	NA	NA
D4: Persistent negative emotional state	NACM	D	NACM	NACM	NA	NA
D5: Loss of interest	NACM	D	NACM	NACM	An	An
D6: Detachment	NACM	D	NACM	NACM	An	An
D7: Restricted affect	NACM	D	NACM	NACM	An	An
E1: Irritability or anger	A	D	DA	EB	DA	EB
E2: Reckless or self-destructive behaviour	A	A	DA	EB	DA	EB
E3: Hypervigilance	A	A	AA	AA	AA	AA
E4: Exaggerated startle response	A	A	AA	AA	AA	AA
E5: Difficulty concentrating	A	D	DA	DA	DA	DA
E6: Sleep disturbance	A	D	DA	DA	DA	DA

A = alterations in arousal and reactivity; AA = anxious arousal; An = anhedonia; AV = avoidance; D = dysphoria; DA = dysphoric arousal; EB = externalizing behaviours; NA = negative affect; NACM = negative alterations in cognitions and mood; R = re-experiencing.

two four-factor models received the most empirical support: the *Emotional Numbing* model (King, Leskin, King, & Weathers, 1998), which divides the DSM-IV avoidance and numbing symptoms into two different factors, and the *Dysphoria* model (Simms, Watson, & Doebeling, 2002), which maintains the re-experiencing and avoidance factors of the Emotional Numbing model and groups some of the arousal symptoms with the numbing symptoms to create a larger, non-specific dysphoria factor separate from the remaining arousal symptoms. The five-factor *Dysphoric arousal* model (Elhai et al., 2011a) combines the above-mentioned models and maintains the re-experiencing, avoidance, and numbing factors of the Emotional Numbing model, but separates dysphoric arousal from anxious arousal. This latter model was the best fitting model in the confirmatory factor analysis (CFA) studies of DSM-IV PTSD (Armour et al., 2016a). CFA studies of DSM-5 PTSD yielded two six-factor models: the *Anhedonia* model (Liu et al., 2014a) and the *Externalizing behaviour* model (Tsai et al., 2015), which both maintain the difference between dysphoric and anxious arousal symptoms. The Externalizing behaviour model separates the arousal factor in a third factor called externalizing behaviours; the Anhedonia model separates Criterion D symptoms into increased negative affect (general distress) and decreased positive affect (anhedonia). The most recent development combines main aspects of both six-factor models into a *Hybrid* model (Armour et al., 2015). In CFA studies of DSM-5 PTSD, the six-factor Anhedonia and seven-factor Hybrid model appear to be superior to the other factor models of PTSD (Armour, Contractor, Shea, Elhai, & Pietrzak, 2016b; Armour et al., 2015; Shevlin, Hyland, Karatzias, Bisson, & Roberts, 2017; Soberon, Crespo, del Mar Gómez-Gutiérrez, Fernández-Lansac, & Armour, 2016; Weathers et al., 2017; Worthman et al., 2016).

Importantly, the majority of these studies have been conducted with self-reported PTSD symptoms on the PCL-5. As factor structure may be dependent on response format (self-report/interview) and modality (frequency/intensity) (Elhai, Palmieri, Biehn, Frueh, & Magruder, 2010; Palmieri, Weathers, Difede, & King, 2007), and as clinician-administered interviews may be more valid to establish a PTSD diagnosis (Armour et al., 2016a), research on the factor structure of PTSD using clinical interviews is needed. So far only two factor studies with the English CAPS-5 interview have been performed and most support was found for the Hybrid and Anhedonia model of DSM-5 PTSD (Hunt, Chesney, Jorgensen, Schumann, & deRoon-Cassini, 2017; Weathers et al., 2017). A factor study with a German CAPS-5 found evidence for the DSM-5 PTSD model, but other models fitted the data even better (Müller-Engelmann et al., 2018). Until now, this is the only published study on the factor structure of PTSD using a translated CAPS-5 interview, therefore more studies in different linguistic and cultural groups are required.

3. Goals of the present study

In sum, the adaptation of the CAPS for DSM-5 warrants new translations and psychometric evaluations of the interview, including a factor analysis. The aims of the current study were therefore (a) to employ a novel stepped crowd-translation process to derive an accurate and clinically relevant Dutch version of the CAPS-5 and (b) to establish the reliability and factor structure of the Dutch CAPS-5. Based on the results of Weathers et al. (2017), Hunt et al. (2017), and Müller-Engelmann et al. (2018), we expected to find good internal consistency and interrater reliability of

the translated version, and a factor structure supporting the six-factor Anhedonia model and the seven-factor Hybrid model.

4. Development of the Dutch CAPS-5

4.1. Cross-culturally valid translation of diagnostic instruments

Many researchers have stressed the importance of a methodologically sound process to derive an appropriate and cross-culturally valid instrument in different languages (Beaton, Bombardier, Guillemin, & Ferraz, 2000; Brislin, 1970; Brislin, Lonner, & Thorndike, 1973; Jones & Kay, 1992; Maneesriwongul & Dixon, 2004; Sousa & Rojjanasrirat, 2011; Sperber, 2004). According to Beaton et al. (2000) a proper adaptation process at least involves the following steps: multiple translations in the target language (TL), synthesis of the translations, back-translations to the source language (SL), consultation of an expert committee review, a test of the pre-final version, and submission of the documentation to the authors of the original instrument.

Clinical experience with the Dutch CAPS for DSM-IV in the Netherlands showed that many patients and clinicians were struggling with the formal wording of the highly structured interview. The release of the CAPS-5 created the opportunity to move beyond a merely linguistic translation process. In addition to a linguistically accurate translation of the CAPS-5 that would correspond to the official Dutch translation of the DSM-5 ensuring content validity, it was deemed important to create a translation that would optimally fit into clinical and research practice. This led to the development and realization of a novel stepped crowd-translation process

to engender the Dutch CAPS-5, which involved a crowd of Dutch psychotrauma professionals (clinicians as well as researchers) and was based on the principles of a cross-cultural adaptation process (Beaton et al., 2000) as well as the Delphi method (Dalkey, 1969; Dalkey & Helmer, 1963; Hsu & Sandford, 2007) for consensus-building by expert-rounds. This process encompassed six translation steps that were carried out to develop the Dutch version of the CAPS-5 (Figure 1).

4.1.1. Step 1: translation

A first professional translator (with Dutch as native language, over 20 years' experience in English-Dutch translations, and knowledgeable in the field of psychotrauma) made a symmetrical and an asymmetrical translation of the CAPS-5. In the former, the goal is loyalty in meaning and equal familiarity and informality in source and target language. In the latter, emphasis is put on a more strict translation with loyalty to the source language (Werner & Campbel, 1970). Simultaneously, a crowd of 44 psychotrauma professionals made English-to-Dutch translations of 4–5 items of the CAPS-5 each. The psychotrauma professionals were recruited through the Dutch Society for Psychotrauma (NtVP) and personal professional networks of the senior experts in the expert panel (see Step 2). The members of the crowd (86.4% female and mean age 40 years, $SD = 12$) were researchers (38.6%), clinicians (36.4%), or both (25.0%). Mean length of self-reported expertise in the psychotrauma field was nine years ($SD = 7$). Additionally, members self-reported their knowledge level of PTSD; 72.7% indicated 'expert-level', 18.2% 'above-average', and 9.1% indicated their knowledge about PTSD as 'average'. Dutch was the native language for all crowd members
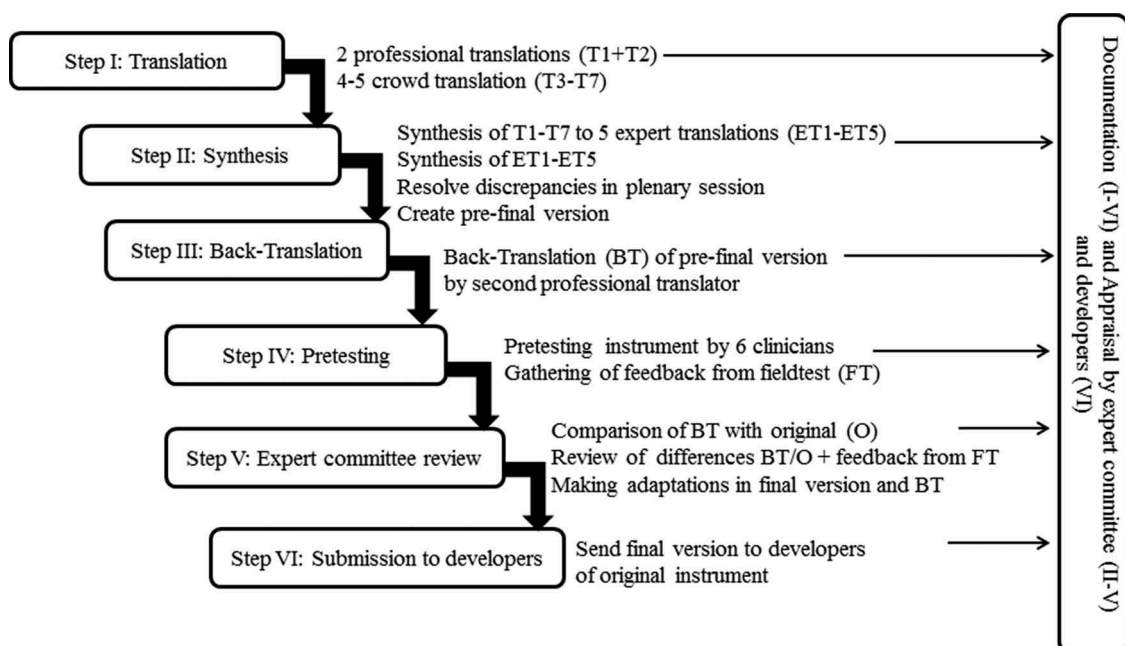

Figure 1. Results per step of the translation project.

and most reported to have an above-average proficiency of the English language (i.e. 70.5%; 29.5% indicated average proficiency). Crowd members were asked to maintain the layout of the CAPS-5 items and propose a correct but foremost a clinically useful translation.

4.1.2. Step 2: synthesis
Each member of an expert panel of five senior psychotrauma experts created a version of the entire CAPS-5. The experts had mixed professional backgrounds (psychiatry, clinical psychology, psychotrauma-related disorders, and anxiety disorders) and either high academic positions or clinical leadership roles. All panel members had above average proficiency of the English language and professional knowledge of PTSD and the CAPS. The expert panel members were instructed to work independently, aggregate the partial crowd-translations based on consensus, maintain the layout of the CAPS-5 and connection to the Dutch DSM-5 criteria with respect to terminology and meaning, use the professional translations as a reference of correctness, and keep the entire CAPS-5 in mind. Thereafter, two project managers (also with above average proficiency of the English language and professional knowledge of PTSD and the CAPS) aggregated the five expert-versions using consensus into one CAPS-5 proposal with 176 significant discrepancies being marked. Discrepancies (either no consensus between expert versions about a core term, or no consensus about a non-core term and both professional translations differed) were marked and presented to the expert panel. During a plenary session with the expert panel consensus was reached about discrepancies, which resulted in a preliminary version of the Dutch CAPS-5.

4.1.3. Step 3: back-translation
The pre-final version was back-translated by an independent second professional translator (with Dutch as native language, eight years' experience in Dutch–English translations and knowledgeable in the field of psychotrauma).

4.1.4. Step 4: pretesting
Six clinicians field-tested the pre-final version, which resulted in multiple questions about altered content of CAPS-5 and a collection of wording suggestions that could facilitate clinical assessment and patient understanding.

4.1.5. Step 5: expert committee review
An independent psychotrauma expert (senior clinician and researcher) with excellent proficiency of the English language and professional knowledge of both PTSD and the CAPS compared the back-translation to the original English CAPS-5 and marked multiple significant differences. These differences and the

suggestions of the field-testers were presented in a final expert committee review, resulting in 45 wording adaptations; the final Dutch CAPS-5 (Boeschoten et al., 2014). The second professional translator made final adaptations in the back-translation.

4.1.6. Step 6: submission of documentation to the developers
The final version including back-translation was sent to the authors of the original CAPS-5 at the National Center for PTSD.

5. Psychometric properties of the Dutch CAPS-5

5.1. Methods

5.1.1. Procedure
From September 2015 onwards, the CAPS-5 was administered as part of a routine clinical assessment to all clients who applied for clinical assessment and/or treatment at Foundation Centrum '45 or Psychotrauma Diagnosis Center (PDC) in the Netherlands. Both centres are partners in the Arq Psychotrauma Expert Group and are specialized in diagnostic assessment of (complex) psychotrauma-related complaints; Foundation Centrum '45 is also specialized in treatment of such complaints. Participants were informed that their data were anonymously used for research purposes. Following a consultation with the medical ethics committee of the Leiden University Medical Center in the Netherlands, no informed consent was asked because the CAPS-5 was primarily conducted for diagnostic purposes and only secondarily for research purposes.

5.1.2. Participants
CAPS-5 data were available for 732 Dutch participants. Twenty-seven (3.7%) participants were excluded because of the absence of a clinician- or self-identified PTSD criterion A event (exposure to actual or threatened death, serious injury, or sexual violence) according to DSM-5. Thirty-six (4.9%) participants were excluded because of incomplete data with regard to the CAPS-5. The total sample that was used for data analyses consisted of 669 Dutch participants with complete data on the CAPS-5: 493 men (73.7%) and 176 women (26.3%). Age at assessment ranged between 21 and 82 years with a mean age of 45.9 years ($SD = 11.7$). The sample consisted of groups with different trauma backgrounds. The largest group (70.1%; $n = 469$) consisted of participants with a profession-related trauma background, mainly first responders such as police officers and fire-fighters. The second largest group (12.7%; $n = 85$) consisted of military veterans, followed by a group (6.0%; $n = 40$) of trauma-exposed adult children of parents traumatized in World War II. The remaining participants had other (11.2%; $n = 75$) trauma backgrounds.

For the evaluation of interrater-reliability, a sample of 35 audio-recorded CAPS-5 interviews was checked for audibility and completeness (i.e. consisting of criterion A–G). Thirty audio recorded interviews appeared to be properly audible and complete. Patients in 20 interviews met a PTSD diagnosis as assessed with CAPS-5, whereas patients in the remaining 10 did not. Ten interviews were randomly selected out of the 20 interviews with patients meeting a PTSD diagnosis in order to obtain a sample of 10 patients meeting a PTSD diagnosis and 10 patients not meeting a PTSD diagnosis. Initial data analyses on double-rated interviews revealed that due to missing data there was insufficient information either to compute a PTSD severity score or to establish a PTSD diagnosis in the second rating of nine interviews (four interviews with patients meeting a PTSD diagnosis and five with patients not meeting a PTSD diagnosis). In order to obtain the targeted sample of 20 double-rated interviews, an additional sample of 29 audio-recorded CAPS-5 interviews was checked. Fifteen of these interviews appeared to be properly audible and complete. From these interviews, four interviews with patients meeting a PTSD diagnosis and five not meeting a PTSD diagnosis were randomly selected. This resulted in a total sample of 29 double-rated interviews. In 20 interviews there was sufficient information from both ratings to compute a PTSD severity score and in 25 interviews there was sufficient information from both ratings to establish a PTSD diagnosis. All CAPS-5 interviews that were used for the first rating were conducted, audio-recorded, and rated by highly experienced CAPS assessors who received initial training by one of the researchers of the research group that developed the original CAPS-5. To obtain second ratings, the audio-recorded interviews were randomly assigned to six experienced CAPS assessors for a second rating who were blind to the original rating. The assessors of the second rating were either trained by one of the researchers of the research group that developed the original CAPS-5, or received training and supervision by an expert CAPS-5 assessor and trainer originally trained by one of the researchers of the group that developed the original CAPS-5.

5.1.3. Diagnostic instruments

The CAPS-5 (Weathers et al., 2013a) is a 30-item interview to assess DSM-5 criteria of PTSD. The newly developed Dutch version of the CAPS-5 (Boeschoten et al., 2014) was used for the purpose of the present study. In the CAPS-5 it is first established whether and how a respondent is exposed to actual or threatened death, serious injury, or sexual violence (Criterion A). Subsequently, 20 PTSD symptoms regarding intrusions (Criterion B), avoidance (Criterion C), negative alterations in cognitions and mood (Criterion D), and alterations in arousal and reactivity (Criterion E) are assessed, as well as three

questions regarding whether the symptoms cause clinically significant distress or impairment (Criterion G). These items are rated on a 5-point severity scale ranging from 0 (*absent*) to 4 (*extreme/incapacitating*). In addition, the CAPS-5 consists of questions regarding the onset and duration of the disturbance (Criterion F) and dissociative symptoms of depersonalization and derealization. With these latter two items a dissociative subtype of PTSD can be determined. In view of the limited research on the factor structure and reliability of the dissociative subtype of PTSD, this subtype and its symptoms were considered outside the scope of the present study.

According to the basic CAPS-5 symptom scoring rule (SEV2 rule), a symptom or impairment is considered present if its severity is rated with 2 or higher (Weathers et al., 2017). Using the DSM-5 algorithm in combination with the SEV2 rule it was established whether or not participants met a diagnosis of PTSD. In addition, by summing the 20 symptom severity scores (Criteria B–E), a total PTSD symptom severity score was computed ranging between 0 and 80 with higher scores indicating higher PTSD symptom severity.

5.1.4. Data analyses

IBM SPSS (Version 23) was used to investigate internal consistency and interrater reliability of the CAPS-5. Internal consistency was evaluated with Cronbach's alpha, inter-item correlations, and item-total correlations. Internal consistency can be considered good when Cronbach's alpha for the total scale is > .80 and most inter-item correlations are in the recommended range of moderate magnitude of .15–.50 (Clark & Watson, 1995). Because Cronbach's alpha is a function of scale length, it was expected to be lower for the symptom clusters Re-experiencing (5 items), Negative alterations in cognitions and mood (7 items), Arousal (6 items), and especially Avoidance (2 items) compared to the total scale. Corrected item-total correlations were computed to assess whether scores on the individual items are associated with the overall PTSD score on the CAPS-5.

Interrater reliability of the categorical PTSD diagnosis was evaluated by calculating Fleiss's kappa (Fleiss, 1971), which was used because the first and second rating of the CAPS-5 interviews were done by more than two different raters. Interrater reliability of a categorical scale is considered good when the kappa coefficient is > .8, substantial when it is > .6, and moderate when it is > .4 (Landis & Koch, 1977). Interrater reliability for the continuous total PTSD symptom severity score was evaluated by calculating a single score intraclass correlation coefficient (ICC), using an absolute agreement definition with a two-way random effects model (McGraw & Wong, 1996). Interrater reliability of a continuous scale is generally defined as good when ICC > .8.

MPlus Version 8 (Muthén & Muthén, 1998–2012) was used to conduct CFA to test and compare six different factor models of PTSD (see Table 1.). Because the symptom severity scores were categorical and skewed, the data were treated as ordinal and the CFAs were estimated with the means and variance adjusted weighted least square (WLSMV) estimator using the THETA parameterization. The two highest response categories were merged into one because of low and zero response frequencies (see supplementary Table 1) with regard to the highest response category for most symptoms. An underlying normal distribution was assumed for each symptom, where the four response categories were divided by three thresholds which were estimated from the data. Comparative Fit Index (CFI), Tucker Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA) were used to assess model fit. For CFI and TLI, model fit is considered good when values are close to .95 (Hu & Bentler, 1999). RMSEA is considered adequate when the value is < .08 and good when it is < .06 (Hu & Bentler, 1999). The difference in goodness-of-fit between nested PTSD models was evaluated by the χ^2 difference test. The 'difftest' option in MPlus was used for appropriate χ^2 difference testing with the WLSMV estimator (Muthén & Muthén, 1998–2012). Some models were however non-nested and it is impossible to compare non-nested models by χ^2 difference testing. The χ^2 difference test is also highly sensitive to sample size such that even trivial differences between two nested models may be significant (Cheung & Rensvold, 2002). Therefore, the Bayesian Information Criterion (BIC) was also used to compare both nested and non-nested models. BIC makes a trade-off between model fit and model complexity and a lower value of BIC is an indication of a better trade-off between model fit and complexity (Van de Schoot, Lugtig, & Hox, 2012). A difference in BIC of 10 points between two models indicates that the model with the lowest BIC value fits the data best (Kass & Raftery, 1995). Because the BIC cannot be obtained by using the WLSMV estimator, all models were re-estimated using the robust maximum likelihood estimator (MLR) to obtain BIC.

5.2. Results

In the total sample, 67.9% (*n* = 442) of the participants met a PTSD diagnosis according to DSM-5. In the total sample, PTSD symptom severity ranged between 0 and 60 with a mean severity level of 29.27 (*SD* = 13.63). Among participants meeting a PTSD diagnosis, PTSD symptom severity ranged between 17 and 60 with a mean severity level of 36.51 (*SD* = 9.06). Descriptive statistics with regard to the items of the CAPS-5 are presented in supplementary Table 1. A total of 594 (90.1%) participants reported an index event that was life-threatening to self (21.8%), other (43.9%), or both

(34.4%). A total of 437 (66.5%) participants reported an index event involving serious injury of self (14.3%), other (72.2%), or both (13.6%). A total of 113 (17.1%) participants reported an index event involving sexual violence affecting self (44.1%), other (46.8%), or both (9.0%).

5.2.1. Internal consistency
Cronbach's alpha coefficients indicated high internal consistency for the full PTSD scale of the CAPS-5 (α = .90) and an acceptable level of internal consistency for the symptom clusters Re-experiencing (α = .74), Avoidance (α = .63), Negative alterations in cognitions and mood (α = .79), and Arousal (α = .72).

Inter-item correlations ranged between .05 and .58 with a mean of .30; 85.3% of inter-item correlations were in the recommended range of moderate magnitude of .15–.50 indicating that, in general, the CAPS-5 features high internal consistency in combination with a differentiated item set. Item D1 (Memory impairment/dissociative amnesia) and E2 (Reckless and self-destructive behaviour) had lower inter-item correlations. Mean inter-item correlation was .12 (range: .05–.21) for item D1 and .17 (range: .08–.25) for item E2. Both items also showed low symptom endorsement (for 17.8 and 10.9% of the participants the severity for item D1 and E2, respectively, was rated as moderate, severe, or extreme), indicating that these items represent rare symptoms of PTSD among our patient population. Inter-item correlations of the remaining 18 items ranged between .14 and .58 with a mean of .34 and 94.8% of the inter-item correlations falling within the recommended range.

Corrected item-total correlations ranged between .19 and .69 with a mean of .52, indicating that, in general, high scores on the CAPS-5 items are associated with high scores on the overall PTSD score of the CAPS-5. Item D1 and E2 showed lower item-total correlations compared to the other items: .19 and .29, respectively. The corrected item-total correlations for the remaining 18 items ranged between .38 and .70 with a mean of .55.

5.2.2. Interrater reliability
Interrater reliability of the total PTSD symptom severity score was high (ICC = .98, 95% CI: .94–.99). There was agreement between raters about the presence of a PTSD diagnosis in 12 (48.0%) interviews and about the absence of a PTSD diagnosis in eight (32.0%) interviews. Raters disagreed about the presence of a PTSD diagnosis in five (20.0%) interviews. Fleiss kappa coefficient was .59 (95% CI: .20–.98), indicating that interrater reliability of the PTSD diagnosis was moderate. To examine when raters tended to disagree about the presence of a PTSD diagnosis, the sample was divided into tertiles (i.e. three groups of equal size divided by the 33rd and 66th percentile) based on the total PTSD symptom

severity score. PTSD symptom severity scores in the first tertile ranged between 2 and 18, scores in the second tertile ranged between 19 and 32, and scores in the third tertile ranged between 33 and 47. All PTSD symptom severity scores resulting from interviews in which there was disagreement about the presence of a PTSD diagnosis fell into the second tertile, whereas only 20% of the PTSD symptom severity scores resulting from interviews in which there was agreement about the presence or absence of a PTSD diagnosis fell into the second tertile (40% fell into both the first and third tertile). This indicates that raters disagreed more often about the presence of a PTSD diagnosis in case of relatively moderate PTSD symptoms, that they agreed more often about the presence or absence of a PTSD diagnosis in case of either relatively mild or more severe PTSD symptoms, and also explains the high interrater reliability of the total PTSD symptom severity in combination with the moderate interrater reliability of the PTSD diagnosis (i.e. participants with moderate PTSD symptom severity scores are frequently at the borderline of the required number of symptoms to meet a PTSD diagnosis). In four out of five cases in which raters disagreed about the presence of a PTSD diagnosis this was due to disagreement about the required number of symptoms regarding negative alterations in cognitions and mood (Criterion D). Disagreement about symptom D4 (Persistent negative emotional state) was involved in three of these cases, and disagreement about symptom D5 (Markedly diminished interest or participation in significant activities) and D7 (Persistent inability to experience positive emotions) was involved in one of these cases. In one out of five cases in which raters disagreed about the presence of a PTSD diagnosis this was due to disagreement about the required number of symptoms regarding avoidance (Criterion C), which involved disagreement in the presence of symptom C1 (Avoidance of or efforts to avoid distressing memories, thoughts, or feelings about or closely associated with the traumatic event). In none of the cases, disagreement about the presence of a PTSD diagnosis was due to disagreement about symptoms regarding intrusions (Criterion B), alterations in arousal and reactivity (Criterion E), or clinically significant distress or impairment (Criterion G).

5.2.3. Confirmatory factor analyses

CFA was used to evaluate and compare six different PTSD models, including two four-factor models (DSM-5 and Dysphoria model), the five-factor Dysphoric arousal model, two six-factor models (Externalizing behaviour and Anhedonia model), and the seven-factor Hybrid model (see respective factors and symptom mappings in Table 1). Table 2 presents model fitting results of these CFAs. CFI, TLI, and RMSEA indicated good model fit for all six PTSD models. The DSM-5 model of PTSD showed a better fit compared to the non-nested Dysphoria model (ΔBIC > 10). The DSM-5 model provided a significantly worse fit compared to the nested Dysphoric arousal, Externalizing behaviour, Anhedonia, and Hybrid models as indicated by the chi-square difference test. This was supported by BIC, except that the more parsimonious DSM-5 model did not fit the data worse compared to the Externalizing behaviour model (ΔBIC < 10). The Dysphoria model fit the data worse compared to all other models. According to the chi-square difference test, the Dysphoric arousal model fit the data significantly worse compared to the nested Externalizing behaviours, Anhedonia, and Hybrid model as indicated by the chi-square difference test. This was supported by BIC, except that the more parsimonious Dysphoric arousal model did not fit the data worse compared to the Externalizing

Table 2. Model fitting results for confirmatory factor analysis of six PTSD models.

Model	χ^2	df	CFI	TLI	RMSEA	BIC	vs.	$\Delta\chi^2$	Δdf	p
							Chi-square difference testing			
1. DSM-5 PTSD model	517.228	164	.964	.958	.057	27,948.565	–	–	–	–
2. Dysphoria	549.065	164	.960	.954	.059	27,959.864	1	–[a]	–[a]	–[a]
3. Dysphoric arousal	489.431	160	.966	.960	.055	27,936.580	1	29.635	4	< .001
							2	52.560	4	< .001
4. Externalizing behaviours	462.519	155	.968	.961	.054	27,943.642	1	57.649	9	< .001
							2	82.424	9	< .001
							3	28.555	5	< .001
5. Anhedonia	**411.853**	**155**	**.973**	**.968**	**.050**	**27,901.143**	**1**	**97.566**	**9**	**< .001**
							2	**120.169**	**9**	**< .001**
							3	**68.681**	**5**	**< .001**
							4	**–[a]**	**–[a]**	**–[a]**
6. Hybrid	378.536	149	.976	.970	.048	27,913.089	1	131.711	15	< .001
							2	156.326	15	< .001
							3	103.963	11	< .001
							4	75.849	6	< .001
							5	34.234	6	< .001

[a] Not applicable, model is not nested in the comparison model; Best fitting model is printed in bold; χ^2, df = chi-square test statistic and degrees of freedom for model; $\Delta\chi^2$, Δdf = chi-square test statistic and degrees of freedom for chi-square difference test between two nested models; vs. = comparison model.

behaviours model (ΔBIC < 10). The Externalizing behaviours model provided a worse fit compared to the non-nested Anhedonia model as indicated by BIC. According to the chi-square difference test and BIC, the Externalizing behaviours model also showed a significantly worse fit compared to the nested Hybrid model. According to the chi-square difference test, the Anhedonia model provided a significantly worse fit compared to the nested Hybrid model. The Anhedonia model, however, showed a better fit compared to the Hybrid model as indicated by BIC (ΔBIC > 10). Taken together, the Anhedonia was selected as the best-fitting model. Not only was this model more parsimonious compared to the Hybrid model, but BIC was also preferred over the chi-square difference test because BIC provides a better trade-off between model fit and complexity (Van de Schoot et al., 2012) and the chi-square difference test is highly sensitive to sample size such that even trivial differences between nested models become significant (Cheung & Rensvold, 2002).

Table 3 presents the unstandardized and standardized factor loadings for the best-fitting Anhedonia model. All factor loadings were significant and the size of the factor loadings indicated that all items, except item D1, were good indicators of their factors. The factor loading of item D1 was relatively lower compared to the other factor loadings, indicating that this item was not a strong indicator of the negative affect factor. Item D1 also showed lower inter-item and item-total correlations and symptom endorsement compared to the other items.

6. Discussion

This paper describes the development and first psychometric evaluation of the Dutch CAPS-5. A novel stepped crowd-translation was carried out, which involved developing the Dutch CAPS-5 with multiple experts in the field of psychotrauma and revising it based on feedback from clinicians who used it in real-world settings. By doing so, evidence of content validity was provided and proof of its fit into clinical practice. The resulting Dutch version of the clinical interview showed a high internal consistency for the total severity score and an acceptable-adequate internal consistency for the symptom cluster scales, as well as a high interrater reliability for the total severity score and moderate interrater reliability for PTSD diagnosis. In addition, we found that an Anhedonia model for PTSD fit present data best.

In line with Weathers et al. (2017) and Müller-Engelmann et al. (2018), high internal consistency for the full PTSD scale of the Dutch CAPS-5 was found and acceptable-adequate internal consistency for the symptom cluster scales, with the lowest for the Avoidance cluster. The latter is most likely due to the fact that the cluster consists of only two items and the Cronbach's alpha measure is a function of scale length (see also Müller-Engelmann et al., 2018; Weathers et al., 2017). The Dutch CAPS-5, like the original version, has differentiated items that were moderately inter-correlated. Our results extend those of Weathers et al. (2017) that Memory impairment (dissociative amnesia, D1) and Reckless and self-destructive behaviour (E2) represent rare symptoms of PTSD. These two items also appeared to be poor indicators of the full PTSD scale in present internal consistency data, again in line with the results of Weathers et al. (2017). In addition, Memory impairment demonstrated a lower factor loading compared to other PTSD symptoms, in line with multiple other studies (e.g. Armour et al., 2015; Keane et al., 2014; Palmieri et al., 2007; Weathers et

Table 3. Unstandardized and standardized factor loading of the six-factor Anhedonia model of PTSD.

Item	Factor	Unstandardized factor loading	SE	Standardized factor loading
B1: Intrusive thoughts	R	1.28	0.10	0.79
B2: Nightmares	R	0.66	0.06	0.55
B3: Flashbacks	R	0.75	0.07	0.60
B4: Psychological cue reactivity	R	1.14	0.08	0.75
B5: Physiological cue reactivity	R	1.07	0.08	0.73
C1: Avoidance of thoughts	AV	1.34	0.13	0.80
C2: Avoidance of reminders	AV	0.92	0.07	0.68
D1: Memory impairment	NA	0.32	0.06	0.31
D2: Negative beliefs	NA	1.02	0.08	0.72
D3: Distorted blame	NA	0.62	0.06	0.53
D4: Persistent negative emotional state	NA	2.10	0.32	0.90
D5: Loss of interest	AN	1.47	0.13	0.83
D6: Detachment	AN	1.40	0.12	0.81
D7: Restricted affect	AN	1.37	0.11	0.81
E1: Irritability or anger	DA	0.77	0.07	0.61
E2: Reckless or self-destructive behaviour	DA	0.52	0.07	0.46
E3: Hypervigilance	AA	1.39	0.15	0.81
E4: Exaggerated startle response	AA	1.02	0.09	0.71
E5: Difficulty concentrating	DA	1.03	0.09	0.72
E6: Sleep disturbance	DA	0.88	0.08	0.66

R = re-experiencing; AV = avoidance; NA = negative affect; AN = anhedonia; DA = dysphoric arousal; AA = anxious arousal

al., 2017), also suggesting it may not be a good indicator of cluster D (negative affect) or PTSD. Therefore, it has been suggested it may need to be removed from this cluster in the future (Bovin et al., 2016). On the other hand, present findings may be related to our sample, which predominantly consisted of participants with a profession-related trauma background. Memory impairment (a dissociative symptom) and Reckless and self-destructive behaviour are symptoms which are considered 'complex' (Friedman, 2013) and may be more often seen in individuals traumatized by interpersonal and childhood traumatic events rather than in individuals traumatized in adult professions. The sample in the study of Weathers et al. (2017), Keane et al. (2014), and Palmieri et al. (2007) also consisted of participants with (predominantly) a profession-related trauma-background (veterans and disaster workers). Interestingly, a recent study with predominantly women suffering from PTSD related to childhood sexual and physical abuse demonstrated higher item-total correlations for both Memory impairment and Reckless and self-destructive behaviour (Müller-Engelmann et al., 2018). Further research in more heterogeneous samples needs to be undertaken to resolve this issue.

Interrater reliability with regard to the total severity score was excellent and within the same range as found by Weathers et al. (2017) and Müller-Engelmann et al. (2018). In this study a lower interrater reliability for PTSD diagnosis was found compared to Weathers et al. (2017). A possible explanation might be that the scoring rule to rate symptom severity is less specific in the CAPS-5 in comparison to the CAPS-IV and therefore possibly more difficult and more error prone. In the CAPS-5 there is more room for the judgement of the clinician as the clinician has to combine information about frequency and intensity before making a single severity rating. Some issues that may offer room for errors and disagreement include (1) for frequency the clinician is not 'forced' to choose a single numeric score on a rating scale, (2) for some items the key rating dimensions for severity use the classification of 'number of times per month' (moderate) and 'number of times per week' (severe) instead of only 'per month' as is the case for frequency, and (3) for some items the key rating dimensions use classifications in 'percentage of time ranges'. Especially for cases with PTSD symptoms in the moderate range it is possible that raters disagree more often about diagnosis as these cases are frequently at the borderline of the required number of symptoms to meet a PTSD diagnosis. For cases with either mild or severe symptoms one may expect more agreement. This is indeed what detailed analyses of double ratings revealed. Detailed analysis also revealed that disagreement about diagnosis was predominantly due to disagreement about the presence of symptoms of the new criterion D regarding negative alterations in cognitions and mood. Specifically, disagreement involved 'persistent negative emotional state' (D4) in three of the five cases and 'persistent inability to experience positive emotions' (D7) in one of the cases. D4 and D7 are separated symptoms, which were grouped together under C6 in CAPS-IV and labelled 'restricted range of affect' (Blake et al., 1995; Friedman, 2013). Results suggest that clinicians had differing interpretations of these altered items. Taken together, our data show that the CAPS-5 symptom and diagnosis scoring algorithm did not lead to unambiguous results on a diagnostic level. Regular consultation between colleagues seems therefore recommended to assure a reliable diagnosis, especially when patients show moderate symptoms and almost/just enough symptoms for a PTSD diagnosis.

Factor structure analyses showed that the DSM-5 PTSD model and five other common PTSD factor models all provided adequate fit to the data. The Anhedonia model for PTSD fit our data best, being more parsimonious than the Hybrid model and showing better fit statistics. This is in line with several other recent studies using DSM-5 self-reports (Shevlin et al., 2017; Soberón et al., 2016). Weathers et al. (2017) concluded that the Anhedonia and Hybrid model both provided best fit to their CAPS-5 data and that the Hybrid model was better based on the chi-square test. Unfortunately, they did not report results for the preferred BIC test (Van de Schoot et al., 2012). It can therefore not be excluded that with a BIC test the Anhedonia model would have had equal or better fit statistics compared to the Hybrid model in their data. Interestingly, Hansen et al. (2017) found that the optimal factor structure of DSM-5 PTSD seems to depend on population. In their data the Hybrid model provided the best fit in a sample of university students and a sample of pain patients, while the Anhedonia model provided the best fit in a military sample.

The Anhedonia model draws upon theoretical and empirical evidence that besides re-experiencing and avoidance, anxious and dysphoric arousal are two separate symptom clusters (Arnberg, Michel, & Johannesson, 2014; Witte, Domino, & Weathers, 2015), and negative affect and decreased positive affect or anhedonia are two unique constructs (Cuthbert & Kozak, 2013; Watson, 2005, 2009; Watson, Clark, & Stasik, 2011). Importantly, the six clusters of the Anhedonia model represent a further refinement *within* the clusters of the DSM-5 model for PTSD. As such, we believe that the Anhedonia model supports the broader DSM-5 PTSD model. Furthermore, the subdivision of the four clusters in the DSM-5 PTSD model is based on a large body of evidence that PTSD four-factor models typically fit well (e.g. Armour et al., 2016a; Elhai & Palmieri, 2011b). Also, the process which resulted in the

diagnostic clusters and criteria for PTSD in DSM-5 was empirically based and rigorous (Friedman, 2013). Additionally, alternative models of PTSD cannot be accepted before more external validation has been done with respect to the ability of their factors to explain causes, consequences, and treatment efficiency (Elhai & Palmieri., 2011b; Palmieri et al., 2007). Moreover, adjusting the diagnosis algorithm of PTSD to even more clusters can alter PTSD prevalence rates (e.g. Elhai, Ford, Ruggiero, & Frueh, 2009; Murphy et al., 2017; Shevlin et al., 2017), which would have implications related to compensation and access to treatment. Besides that, recent studies investigating ICD-11 informed factor models suggest that a more simple factor structure, with core symptoms of PTSD, may perhaps better reflect the configuration of PTSD (Hansen et al., 2017; Hunt et al., 2017). Another proposal, different from either the broad DSM-5 or narrow ICD-11 definition of PTSD, is an 'hierarchical model' in which PTSD would be composed of a *higher-order dysphoria factor* with symptoms shared with other mood and anxiety disorders and a *lower-level PTSD factor* with PTSD specific (core) symptoms. This would allow the possibility to examine the aetiology and pathogenesis of PTSD within an integrative model for mood and anxiety disorders (Hunt et al., 2017; Rademaker et al., 2012; Watson, 2005). Ultimately, it seems that more research is needed to determine which PTSD model best captures PTSD symptomatology and is clinically most meaningful.

We found insufficient evidence for the Externalizing behaviours cluster, which probably explains why in our data the Hybrid model for PTSD did not fit significantly better than the Anhedonia model. The Externalizing behaviour factor is based on literature proposing that individuals with PTSD may show aggression, recklessness, and suicidality (Friedman, 2013), which may be indicative of difficulties with emotion regulation and impulse control (Steiner, Garcia, & Matthews, 1997). Indeed Armour et al. (2016b) found that compared to the other PTSD factors the Externalizing behaviours factor was most strongly related to anger and impulsivity. Like us, multiple other studies found inferior fit statistics for the Externalizing behaviour model compared to the Anhedonia and Hybrid model (Armour et al., 2015, 2016b; Shevlin et al., 2017; Weathers et al., 2017; Wortmann et al., 2016; Yang et al., 2017). It may be explained by the finding that one of the symptoms (Reckless or self-destructive behaviour) in this proposed cluster, composed of only one other symptom (Irritability or anger), occurs infrequently among trauma-exposed individuals (Weathers et al., 2017). However, again it may be related to trauma background of the sample, as recently in a somewhat different sample of Malaysian adolescents, the largest percentage of whom were directly or indirectly exposed to a natural

disaster and/or reported experiencing a transportation accident, the Externalizing behaviour and Hybrid model for PTSD were found to be superior compared to the other PTSD factor models (Murphy et al., 2017). Also, Müller-Engelmann et al. (2018) showed that the Externalizing behaviours model was the model with the best fit without methodological problems in their sample of (mostly) women with a history of childhood sexual and physical abuse.

7. Strengths and limitations

This paper describes an innovative approach of translation to ensure cross-cultural validity of diagnostic instruments using a combination of principles of a cross-cultural adaptation process (Beaton et al., 2000) and the Delphi method (Dalkey, 1969; Dalkey & Helmer, 1963; Hsu & Sandford, 2007). By involving multiple experts in different rounds the employed approach of translation was instrumental in achieving an instrument that was broadly supported among experts. This was illustrated by the finding that synthesis of partial crowd-translations by multiple psychotrauma experts into five independent proposals by senior experts resulted in a great amount of consensus. Also, by involving clinicians and undertaking a field test of the preliminary version the present method assisted in providing a clinically useful instrument. Psychometric evaluation was carried out in a large sample of trauma-exposed individuals using clinical interviews administered by highly trained and experienced assessors.

This study also has several limitations. First, although our sample was more mixed in trauma background and consisted of more females compared to Weathers et al. (2017), it still consisted of mostly males and individuals with profession-related traumatization. Second, we did not establish test-retest reliability and congruent or discriminant validity of the Dutch CAPS-5, nor did we test the external validity of the examined factor models.

In a future project the developers of the original instrument could be involved in the expert panel, to achieve maximal linguistic agreement between the instrument in target and source language. More extensive field testing, in which patient understanding is assessed for each part of the instrument, could further enhance clinical usability (Sousa & Rojjanasrirat, 2011). A survey among the expert crowd inquiring about their rating of or experiences with the preliminary version(s) of the CAPS-5 could further increase support for the translated version by clinicians and researchers in the field of psychotrauma.

Further research is needed into the validity of the (Dutch) CAPS-5 and the factors of the Anhedonia and alternative (more simple or hierarchical) models, including other clinical interviews or self-report

measures, behavioural observations, and measures of etiologic and clinical significance (Armour et al., 2016a; Elhai & Palmieri, 2011b; Murphy et al., 2017; Shevlin et al., 2017; Weathers et al., 2017). It is also important that the psychometric properties of the CAPS-5 and current factor structure results are repeated in different types of samples, differing in trauma background and sociodemographic variables such as gender and cultural background (Elhai & Palmieri, 2011b).

8. Conclusion

In conclusion, this study provides evidence that the Dutch version of the CAPS-5 is a carefully translated instrument with overall adequate psychometric properties. Also, current results add to the growing support for more refined (six and seven) factor models for DSM-5 PTSD indicating that the validity and clinical implications of these models should be objective of further research.

ORCID

Anne Bakker ⓘ http://orcid.org/0000-0002-3248-2441

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Armour, C., Ross, J., & Elhai, J. (2016a). A systematic literature review of PTSD's latent structure in the diagnostic and statistical manual of mental disorders: DSM-IV to DSM-5. *Clinical Psychology Review*, 44, 60–74.

Armour, C., Contractor, A., Shea, T., Elhai, J. D., & Pietrzak, R. H. (2016b). Factor structure of the PTSD checklist for DSM-5: Relationships among symptom clusters, anger, and impulsivity. *Journal of Nervous Mental Disorder*, 204, 108–115.

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.

Arnberg, F. K., Michel, P. O., & Johannesson, K. B. (2014). Properties of Swedish posttraumatic stress measures after a disaster. *Journal of Anxiety Disorders*, 28, 402–409.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. S. (2000). Guidelines for the process of cross-cultural adaption of self-report measures. *Spine*, 25, 3186–3191.

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

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.

Boeschoten, M. A., Bakker, A., Jongedijk, R. A., Van Minnen, A., Elzinga, B. M., Rademaker, A. R., & Olff, M. (2014). *Clinician-administered PTSD scale for DSM-5 Nederlandstalige versie [Dutch version]*. Diemen, The Netherlands: Arq Psychotrauma Expert Group. Retrieved from http://www.psychotraumadiagnostics.centrum45.nl/en

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.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185–216.

Brislin, R. W., Lonner, W., & Thorndike, R. M. (1973). *Cross-cultural research methods*. New York: John Wiley & Sons.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit-indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319.

Cuthbert, B. N., & Kozak, M. J. (2013). Constructing constructs for psychopathology: The NIMH research domain criteria. *Journal of Abnormal Psychology*, 122, 928–937.

Dalkey, N. C. (1969). An experimental study of group opinion. *Futures*, 1, 408–426.

Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9, 458–467.

Elhai, J., Palmieri, P. A., Biehn, T. L., Frueh, B. C., & Magruder, K. M. (2010). Posttraumatic stress disorder's frequency and intensity ratings are associated with factor structure differences in military veterans. *Psychological Assessment*, 22, 723–728.

Elhai, J. D., Biehn, T. L., Armour, C., Klopper, J. J., Frueh, B. C., & Palmieri, P. A. (2011a). Evidence for a unique PTSD construct represented by PTSD's D1–D3 symptoms. *Journal of Anxiety Disorders*, *25*, 340–345.

Elhai, J. D., Ford, J. D., Ruggiero, K. J., & Frueh, B. C. (2009). Diagnostic alterations for post-traumatic stress disorder: Examining data from the national comorbidity survey replication and national survey of adolescents. *Psychological Medicine*, *39*, 1957–1966.

Elhai, J. D., & Palmieri, P. A. (2011b). The factor structure of posttraumatic stress disorder: A literature update, critique of methodology, and agenda for future research. *Journal of Anxiety Disorders*, *25*, 849–854. Review.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *88*, 378–382.

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.

Hansen, M., Hyland, P., Karstoft, K. I., Vaegter, H., Bramsen, R., Nielsen, A., … Andersen, T. (2017). Does size really matter? A multisite study assessing the latent structure of the proposed ICD-11 and the DSM-5 diagnostic criteria for PTSD. *European Journal of Psychotraumatology*, *8*. doi:10.1080/20008198.2017.1398002

Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research and Evaluation*, *12*, 1–8.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Hunt, J. C., Chesney, S. A., Jorgensen, T. D., Schumann, N. R., & deRoon-Cassini, T. A. (2017). Exploring the gold-standard: Evidence for a two-factor model of the clinician administered PTSD scale for the DSM-5. *Psychological Trauma: Theory, Research, Practice, and Policy, Advance Online Publication.* doi:10.1037/tra0000310

Jones, E. G., & Kay, M. (1992). Instrumentation in cross-cultural research. *Nursing Research*, *41*, 186–188.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

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 Assesssment*, *26*, 1138–1145.

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

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Liu, P., Wang, L., Cao, C., Wang, R., Zhang, J., Zhang, B., … Elhai, J. D. (2014a). The underlying dimensions of DSM-5 posttraumatic stress disorder symptoms in an epidemiological sample of Chinese earthquake survivors. *Journal of Anxiety Disorders*, *28*, 345–351.

Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A methods review. *Journal of Advanced Nursing*, *48*, 175–186.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46.

Müller-Engelmann, M., Schnyder, U., Dittmann, C., Priebe, K., Bohus, M., Thome, J., & Steil, R. (2018). Psychometric properties and factor structure of the German version of the clinician-administered PTSD scale for DSM-5. *Assessment*, 1–11. doi:10.1177/1073191118774840

Murphy, S., Hansen, M., Elklit, A., Yong Chen, Y., Raudzah Ghazali, S., & Shevlin, M. (2017). Alternative models of DSM-5 PTSD: Examining diagnostic implications. *Psychiatry Research*, *S0165–S01781*, 31234–31239.

Muthén, L. K., & Muthén, B. O. (1998–2012). *MPlus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Olff, M. (2015). Choosing the right instruments for psychotrauma related research. *European Journal of Psychotraumatology*, *6*, 30585. doi:10.3402/ejpt.v6.30585

Palmieri, P. A., Weathers, F. W., Difede, J., & King, D. W. (2007). Confirmatory factor analysis of the PTSD checklist and the clinician-administered PTSD Scale in disaster workers exposed to the world trade center ground zero. *Journal of Abnormal Psychology*, *116*, 329–341.

Rademaker, A., Minnen, A., Ebberink, F., Zuiden, M., Hagenaars, M., & Geuze, E. (2012). Symptom structure of PTSD: Support for a hierarchical model separating core PTSD symptoms from dysphoria. *European Journal of Psychotraumatology*, *3*, 17580.

Shevlin, M., Hyland, P., Karatzias, T., Bisson, J. I., & Roberts, N. P. (2017). Examining the disconnect between psychometric models and clinical reality of posttraumatic stress disorder. *Journal of Anxiety Disorders*, *47*, 54–59.

Simms, L. J., Watson, D., & Doebbeling, B. N. (2002). Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. *Journal of Abnormal Psychology*, *111*, 637–647.

Soberón, C., Crespo, M., Del Mar Gómez-Gutiérrez, M., Fernández-Lansac, V., & Armour, C. (2016). Dimensional structure of DSM-5 posttraumatic stress symptoms in Spanish trauma victims. *European Journal of Psychotraumatology*, *7*, 32078.

Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, *17*, 268–274.

Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology*, *126*(Supplement 1), S124–S128.

Steiner, H., Garcia, I. G., & Matthews, Z. (1997). Posttraumatic stress disorder in incarcerated juvenile delinquents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*, 357–365.

Tsai, J., Harpaz-Rotem, I., Armour, C., Southwick, S. M., Krystal, J. H., & Pietrzak, R. H. (2015). Dimensional structure of DSM-5 posttraumatic stress disorder symptoms: Results from the National Health and Resilience in Veterans Study. *The Journal of Clinical Psychiatry*, *76*, 546–553.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492.

Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology*, *114*, 522–536.

Watson, D. (2009). Differentiating the mood and anxiety disorders: A quadripartite model. *Annual Review of Clinical Psychology*, *5*, 221–247.

Watson, D., Clark, L. A., & Stasik, S. M. (2011). Emotions and the emotional disorders: A quantitative hierarchical

perspective. *International Journal of Clinical and Health Psychology*, 11, 429–442.

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013a). *Clinician-administered PTSD scale for DSM-5 (CAPS-5)*. Boston, MA: National Center for PTSD.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013b). *The PTSD checklist for DSM-5 (PCL-5)*. Boston, MA: National Center for PTSD.

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., … Marx, B. P. (2017). The clinician-administered PTSD scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*. Advance online publication. doi:10.1037/pas0000486

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.

Werner, O., & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of method in cultural anthropology* (pp. 398–420). New York: Columbia University Press.

Witte, T. K., Domino, J. L., & Weathers, F. W. (2015). Item order effects in the evaluation of posttraumatic stress disorder symptom structure. *Psychological Assessment*, 27, 852–864.

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.

Yang, H., Wang, L., Cao, C., Cao, X., Fang, R., Zhang, J., & Elhai, J. D. (2017). The underlying dimensions of DSM-5 PTSD symptoms and their relations with anxiety and depression in a sample of adolescents exposed to an explosion accident. *European Journal of Psychotraumatology*, 8, 1272789.