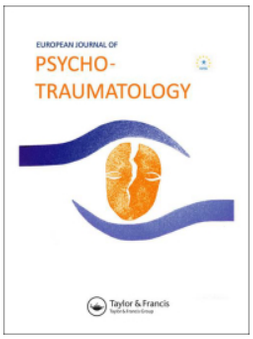

European Journal of Psychotraumatology

ISSN: 2000-8066 (Online) Journal homepage: www.tandfonline.com/journals/zept20

Crossing cultural barriers: an initial cross-cultural validation of the Arabic compared to the German version of the Posttraumatic Stress Disorder Checklist for DSM-5 using multi-group confirmatory factor analysis

Charina C. Lüder, Monika Equit, Nicolas Becker, Antje Schönfelder, Heide Glaesmer, Yuriy Nesterko & Tanja Michael

To cite this article: Charina C. Lüder, Monika Equit, Nicolas Becker, Antje Schönfelder, Heide Glaesmer, Yuriy Nesterko & Tanja Michael (2025) Crossing cultural barriers: an initial cross-cultural validation of the Arabic compared to the German version of the Posttraumatic Stress Disorder Checklist for DSM-5 using multi-group confirmatory factor analysis, European Journal of Psychotraumatology, 16:1, 2565898, DOI: 10.1080/20008066.2025.2565898

To link to this article: https://doi.org/10.1080/20008066.2025.2565898

EUROPEAN JOURNAL OF PSYCHOTRAUMATOLOGY
2025, VOL. 16, NO. 1, 2565898
https://doi.org/10.1080/20008066.2025.2565898

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

Taylor & Francis
Taylor & Francis Group

CLINICAL RESEARCH ARTICLE

OPEN ACCESS Check for updates

Crossing cultural barriers: an initial cross-cultural validation of the Arabic compared to the German version of the Posttraumatic Stress Disorder Checklist for DSM-5 using multi-group confirmatory factor analysis

Charina C. Lüder [a], Monika Equit [a], Nicolas Becker[b], Antje Schönfelder [c], Heide Glaesmer [c], Yuriy Nesterko [c,d,e] and Tanja Michael [a]

[a]Division of Clinical Psychology and Psychotherapy, Department of Psychology, Saarland University, Saarbrücken, Germany; [b]Division of Differential Psychology and Psychological Assessment, Department of Psychology, Saarland University, Saarbrücken, Germany; [c]Department of Medical Psychology and Medical Sociology, University Leipzig, Leipzig, Germany; [d]Department for Clinical Psychological Intervention, Freie Universität Berlin, Berlin, Germany; [e]Department for Traumatic Stress and Transcultural Studies, Center ÜBERLEBEN, Berlin, Germany

ABSTRACT

Background: Post-traumatic stress disorder (PTSD) is prevalent worldwide, yet its phenomenology and prevalence vary according to individual and contextual factors. Due to heightened exposure to (post-) conflict environments, many Arabic-speaking individuals are at high risk of PTSD. The PTSD Checklist for DSM-5 (PCL-5) is a widely used screening tool for PTSD symptoms, validated in several languages, including German and Arabic. However, despite its frequent cross-linguistic and cross-cultural use, a comprehensive cross-linguistic and cross-cultural validation of the PCL-5 Arabic version still remains outstanding.

Objective: To ensure the cross-linguistic and cross-cultural comparability of the PCL-5 German and Arabic versions, this study examined the measurement invariance in a heterogeneous sample of German-speaking ($n = 283$) and Arabic-speaking individuals ($n = 295$).

Method: Sociodemographic data and characteristics of stressful life events were assessed. Subsequently, we examined the internal consistency of the PCL-5 Arabic and German versions and broaden current investigations on structural validity as conducted via confirmatory factor analyses (CFA) by multi-group confirmatory factor analyses (MGCFA) across both language versions.

Results: The present findings show that the Arabic-speaking subsample reported more man-made trauma, which was associated with higher PCL-5 sum scores compared to the German-speaking subsample. The PCL-5 showed excellent internal consistency (Cronbach's $\alpha = .96$). CFA indicated good model fit for all models tested, favouring the Anhedonia and Hybrid models. While MGCFA confirmed configural, threshold, and metric invariance, the scalar invariance could not be established.

Conclusions: The present study supports previous research indicating that the factorial structure of the PCL-5 is consistent across both language versions in the CFA. Nevertheless, our findings show a lack of scalar invariance in the MGCFA, which suggests potential bias in the cross-linguistic and cross-cultural comparability of the PCL-5 sum scores between the Arabic and the German versions. This highlights the need for context-, language-, and culture-sensitive diagnostics to ensure accurate PTSD assessments.

Cruzando barreras culturales: una validación transcultural inicial de la versión árabe en comparación con la versión alemana de la Lista de Verificación para el Trastorno de Estrés Postraumático según el DSM-5 mediante un análisis factorial confirmatorio multigrupo

Antecedentes: El trastorno de estrés postraumático (TEPT) es prevalente en todo el mundo, aunque su fenomenología y prevalencia varían según factores individuales y contextuales. Debido a la elevada exposición a entornos de conflicto y posconflicto, muchas personas de habla árabe presentan un alto riesgo de desarrollar TEPT. La Lista de Chequeo del TEPT según el DSM-5 (PCL-5, según sus siglas en inglés) es una herramienta de detección de síntomas de TEPT ampliamente utilizada, validada en varios idiomas, incluidos el alemán y el árabe. Sin embargo, a pesar de su frecuente uso translingüístico y transcultural, aún queda pendiente una validación translingüística y transcultural integral de la versión árabe del PCL-5.

Objetivo: Para garantizar la comparabilidad translingüística y transcultural de las versiones alemana y árabe de la PCL-5, este estudio examinó la invarianza de medición en una

ARTICLE HISTORY

Received 28 March 2025
Revised 8 September 2025
Accepted 17 September 2025

KEYWORDS

Posttraumatic Stress Disorder Checklist for DSM-5; cross-cultural validation; measurement invariance

PALABRAS CLAVES

Lista de Verificación para el Trastorno de Estrés Postraumático según el DSM-5 (PCL-5); validación transcultural; invarianza de medida

HIGHLIGHTS

- This study goes beyond previous research as the first to examine measurement invariance of PCL-5 German and Arabic versions in a direct European–Middle Eastern comparison.
- No significant discrepancies in factor structures emerged when comparing the confirmatory factor analyses of the German and Arabic PCL-5 versions, supporting the best fit for the Hybrid and Anhedonia models.
- The examination of measurement invariance between the PCL-5 German and Arabic versions supports configural, metric, and threshold invariance, highlighting comparable six- to seven-factor structures with similar item-factor correlation

CONTACT Charina C. Lüder charina.lueder@uni-saarland.de Division of Clinical Psychology and Psychotherapy, Department of Psychology, Saarland University, Building A1.3, D-66123 Saarbrücken, Germany

Supplemental data for this article can be accessed online at https://doi.org/10.1080/20008066.2025.2565898

muestra heterogénea de personas germanoparlantes ($n = 283$) y arabo parlantes ($n = 295$).
Método: Se evaluaron los datos sociodemográficos y las características de los acontecimientos vitales estresantes. Posteriormente, examinamos la consistencia interna de las versiones árabe y alemana de la PCL-5, y ampliar las investigaciones actuales sobre la validez estructural, realizadas mediante análisis factorial confirmatorio (CFA, según sus siglas en inglés), a través de análisis factorial confirmatorio multigrupo (MGCFA, según sus siglas en inglés) en ambas versiones lingüísticas.
Resultados: Los resultados actuales muestran que la submuestra arabo parlante reportó más traumas provocados por el ser humano, lo cual se asoció con puntuaciones totales más altas en la PCL-5 en comparación con la submuestra germanoparlante. La PCL-5 mostró una consistencia interna excelente (α de Cronbach = 0,96). El CFA indicó un buen ajuste del modelo para todos los modelos evaluados, favoreciendo los modelos de Anhedonia e Híbrido. Aunque el MGCFA confirmó la invarianza configural, de umbral y métrica, no fue posible establecer la invarianza escalar.
Conclusión: El presente estudio respalda investigaciones previas que indican que la estructura factorial de la PCL-5 es consistente en ambas versiones lingüísticas en el CFA. Sin embargo, nuestros hallazgos muestran ausencia de invarianza escalar en el MGCFA, lo que sugiere un posible sesgo en la comparabilidad translingüística y transcultural de las puntuaciones totales de la PCL-5 entre las versiones árabe y alemana. Esto resalta la necesidad de diagnósticos sensibles al contexto, al idioma y a la cultura, para garantizar evaluaciones precisas del TEPT.

Abbreviations: CAPS-5: Clinican-Administered PTSD Scale for DSM-5; CFA: Confirmatory factor analysis; CFI: Comparative fit index; *DSM-5*: *The Diagnostic and Statistical Manual of Mental Disorders, 5th Edition;* FIML: Full Information Maximum Likelihood; IES-R: Impact of Events Scale-Revised; LEC-5: Life Events Checklist for DSM-5; MGCFA: Multi-group confirmatory factor analysis; PCL-5: Posttraumatic Stress Disorder Checklist for DSM-5; PTSD: Posttraumatic stress disorder; RMSEA: Root mean square error of approximation; ROC: Receiver operating characteristic; SRMR: Standardised root mean square residual; TLI: Tucker–Lewis index; WLSMV: Weighted least square mean and variance.

patterns, whereas missing scalar invariance indicates potential bias in comparing PCL-5 sum scores between these two subsamples.

1. Introduction

Post-traumatic stress disorder (PTSD) is a prevalent psychological response to traumatic experiences (Koenen et al., 2017) and is commonly assessed using PTSD assessments such as the PTSD Checklist for DSM-5 (PCL-5) (Forkus et al., 2022). The development of PTSD assessment tools is predicated on the theoretical conceptualisation of the PTSD symptom structure. Several theoretical models of PTSD symptom structure have been proposed in previous research, with these models being assessed for their applicability across diverse samples.

1.1. Conceptualisation of PTSD

While the DSM-5 (American Psychiatric Association, 2013) proposes a four-factor model of PTSD, including *intrusion, avoidance, negative alterations in cognition and mood*, and *alterations in arousal and reactivity*, current research has indicated limited stability of this model (Armour et al., 2015; Blevins et al., 2015) and suggested that alternative structural models, such as the Dysphoria model, the Dysphoric arousal model, the Anhedonia model, the Externalising behaviour model and the Hybrid model may provide a better model fit. For an overview see Supplementary Materials 1 (SM1). The Dysphoria model (Simms et al., 2002) and the Dysphoric arousal model (Elhai et al., 2011) were derived from the DSM-5 model. The Dysphoria model comprises four factors: *intrusion, avoidance, dysphoria*, and *alterations in arousal and reactivity*. In this model, the factor of *negative alterations in cognitions and mood* was replaced by *dysphoria*, which includes the symptoms of irritability or aggression, difficulties concentrating, and difficulties sleeping. *Dysphoria* refers to a state of general dissatisfaction or discomfort, often associated with strong feelings of sadness, irritability, or anxiety. As a result, the factor of *alterations in arousal and reactivity* includes three symptoms in the Dysphoria model instead of six symptoms, as in the DSM-5 model: risky and destructive behaviour, hypervigilance, and an exaggerated startle response. The Dysphoric arousal model consists of five factors, with the *alterations in arousal and reactivity* from the DSM-5 model separated into two distinct clusters of hyperarousal: *dysphoric arousal* and *anxious arousal.* The Anhedonia (Liu et al., 2014) and Externalising behaviour models (Tsai et al., 2014) were derived from the Dysphoric arousal model and were extended to six-factor models. The Anhedonia model separates *negative alterations in cognition and mood* into two distinct factors: changes in *negative affect* and *anhedonia*. The Externalising behaviour model also comprises six factors. Unlike the Anhedonia model, though, the Externalising behaviour model combines the factors of loss of positive affect (*anhedonia*) and *negative affect* into a single overarching factor entitled *numbing*. Further, *dysphoric arousal* was divided into

two separate factors: *externalising behaviour* and *dysphoric arousal*. Finally, the Hybrid model (Armour et al., 2015) integrates the previously proposed Anhedonia and Externalising behaviour models into a comprehensive seven-factor model, including *intrusion, avoidance, negative affect, anhedonia, externalising behaviour, anxious arousal,* and *dysphoric arousal.*

1.2. Assessment of PTSD: post-traumatic stress disorder checklist for DSM-5

The PCL-5 is a self-report screening tool that assesses the severity of PTSD symptoms based on the DSM-5 criteria for PTSD, referring to a most distressing traumatic event as defined by Criterion A of the DSM-5 (American Psychiatric Association, 2013). Traumatic events are identified using the Life Events Checklist (LEC-5; Weathers et al., 2013a). Both the PCL-5 and LEC-5 are available in multiple languages (Hoffman et al., 2022), including German and Arabic, and are widely used in clinical and research settings across diverse populations (e.g. Lüder et al., 2023), requiring cross-group psychometric robustness of the PCL-5.

1.3. Internal consistency of the PCL-5 in a cross-cultural comparison

Internal consistency is a key aspect of psychometric robustness and a fundamental prerequisite for structural validity, as it initially assesses how reliably the items of the PCL-5 capture PTSD symptoms. In this context, Forkus et al. (2022) reviewed $n = 47$ studies on the psychometric properties of the PCL-5 and concluded, that overall, the PCL-5 is a psychometrically strong assessment tool with good internal consistency (Cronbach's $\alpha > .80$). A psychometric evaluation of the German version of the PCL-5 with a clinical sample also showed high internal consistency ($\alpha = .95$; Krüger-Gottschalk et al., 2017). Ibrahim et al. (2018) found good internal consistency ($\alpha = .85$) for the Arabic version of the PCL-5 in a war-affected sample, while Brahim et al. (2022) reported excellent internal consistency ($\alpha = .98$) for the Tunisian Arabic version in a military sample.

1.4. Structural validity of the post-traumatic stress disorder checklist for DSM-5 in a global and cross-cultural comparison

Relying on the substantial internal consistency of the PCL-5, current research has focused on evaluating its structural validity, an essential step in assessing the extent of which theoretical models of PTSD correspond to its clinical manifestation. The findings obtained to date offer a slightly heterogeneous perspective on which theoretical model most accurately aligns with the data, with studies conducted across diverse samples yielding inconsistent results: Forkus et al. (2022) reviewed findings from 39 studies, most of which conducted factor analyses on the English version of the PCL-5 in White, military, or university samples. In 15 of these studies, the seven-factor Hybrid model was identified as the best-fitting model, while three studies reported comparable fit between the Anhedonia and Hybrid models (e.g. Blevins et al., 2015).

Krüger-Gottschalk et al. (2017) and Pettrich et al. (2024) examined the structural validity of the German version of the PCL-5, with inconclusive results in Krüger-Gottschalk et al. (2017), while Pettrich et al. (2024) identified the Hybrid model as the best fitting model encompassing all PCL-5 items.

Ibrahim et al. (2019) showed findings from the CFA of the Arabic version of the PCL-5 in favour of the Anhedonia and Hybrid models. The authors further extended single-group CFA to multi-group CFA (MGCFA) between the Arabic and Kurdish versions of the PCL-5 in a sample of war-affected displaced persons from Iraq and Syria. The MGCFAs revealed configural, metric, and scalar invariance between the Arabic and Kurdish versions of the PCL-5.

1.5. The global and cross-cultural prevalence of PTSD

Within its widespread application in clinical and research settings, the PCL-5 serves not only as a tool for assessing PTSD symptom severity, but also for estimating prevalence rates of PTSD, taking into account varying cut-off scores. Notably, the estimated lifetime prevalence of PTSD in the general world population is 3.9% (Koenen et al., 2017). However, prevalence of trauma exposure (approximately 70%) far exceeds that of PTSD (Kessler et al., 2017). The global prevalence of PTSD varies considerably, as it is influenced by environmental context, which affects both the frequency (multiple vs. single events) and type (man-made vs. accidental) of trauma. Particularly Arabic-speaking refugees and asylum seekers from North Africa and the Middle East are frequently exposed to severe human rights violations, further exacerbated by ongoing conflict and political instability following the Arab Spring (Ibrahim et al., 2019; Quosh et al., 2013). Consequently, they face a heightened risk of cumulative man-made trauma (Fazel et al., 2005; Nesterko et al., 2020), significantly increasing their vulnerability to PTSD (AlShawi, 2018; Ibrahim et al., 2018; Nesterko et al., 2020).

1.6. Gap in current research

Considering this comprehensive body of research, it can be summarised that despite the higher prevalence

of PTSD in conflict-affected populations from low- and middle-income countries, the factor structure underlying the PCL-5 has predominantly been examined in populations from non-war-affected, high-income countries (Forkus et al., 2022). To date, research has not directly compared the structural validity of the PCL-5 between the Arabic and German versions. The paucity of cross-linguistic and cross-cultural investigations of the PCL-5's factor structure poses significant limitations to its generalizability and applicability across diverse populations.

1.7. Aims of the current study

To address this gap, the present study examines the measurement invariance of the German (Krüger-Gottschalk et al., 2017) and Arabic (Ibrahim et al., 2018) versions of the PCL-5 among German- and Arabic-speaking individuals in Germany. To achieve this, we first assess trauma exposure (measured with the LEC-5) and PTSD symptoms (measured with the PCL-5). Based on previous research (e.g. Georgiadou et al., 2017), we expect the number of traumatic experiences and PTSD symptoms to be higher in the Arabic-speaking subsample compared to the German subsample. To advance cross-cultural research on the structural validity of the PCL-5, subsample and total sample CFAs of the six most common factor models (see SM1) assess structural validity of both versions, guiding subsequent analyses of (configural, threshold, metric, scalar, and full) measurement invariance using MGCFA. In accordance with current research, it is hypothesised that the Anhedonia and Hybrid models will demonstrate the best model fit across both the Arabic and German subsamples, thereby providing evidence for the assessment of a similarly conceptualised construct of PTSD.

2. Methods

2.1. Participants

Following established recommendations (e.g. Chen, 2008 Davidov et al., 2014; Milfont & Fischer, 2010; Putnick & Bornstein, 2016), a target sample size of approximately 300 participants per group was aimed for, as this is considered sufficiently large for testing measurement invariance. A total of $n = 630$ non-clinical and clinical participants were recruited for the data analysis. Participants with incomplete PCL-5 data ($n = 52$, 8.3%) were excluded using listwise deletion, resulting in a final sample of $n = 578$ participants. Participants completed either the German or Arabic version of the PCL-5, resulting in two groups: German-speaking ($n = 295$) and Arabic-speaking ($n = 283$)[1]. In the present study, data on participants' age, gender (Langeland & Olff, 2024), nationality, and country of origin was collected. Moreover, the participants' mental health status was ascertained by the question: 'Are you currently experiencing psychological distress?'. Participants who responded affirmatively to this question were classified as 'clinical' in the absence of a formal diagnosis. These participants were subsequently asked whether they were currently receiving psychiatric or psychotherapeutic treatment. A comprehensive overview of the sample's characteristics is provided in *Table 1*.

2.2. Measures

The questionnaires on sociodemographic data were professionally translated and back-translated in a similar way as is used for the translations of standardised instruments for the assessment of PTSD (see Krüger-Gottschalk et al., 2017; Nesterko et al., 2020)

2.2.1. Exposure to stressful life events

The German (Krüger-Gottschalk et al., 2017) and Arabic (Nesterko et al., 2020) versions of the LEC-5 (Weathers et al., 2013a), a structured 16-item assessment tool, were used to assess criterion A of PTSD symptomatology, namely exposure to traumatic events (American Psychiatric Association, 2013). The LEC-5 has been internationally validated and is available in several languages, making it a frequently referenced instrument for the detailed examination of trauma exposure in various contexts. Specifically, the LEC-5 queries the traumatic event to which the PCL-5 refers. If no traumatic event could be identified on the LEC-5, participants were asked to report their most stressful life event as a reference for answering the PCL-5.

2.2.2. PTSD

PTSD symptoms over the past month were assessed using the PCL-5 (Blevins et al., 2015), a 20-item self-report assessment tool rated on a five-point Likert scale from 0 (*not at all*) to 4 (*extremely*). The scores for each item are aggregated to yield an overall severity sum score ranging from 0 to 80, with higher scores indicating more severe symptoms. As with the LEC-5, we used the German (Krüger-Gottschalk et al., 2017) and Arabic (Nesterko et al., 2020) versions of the PCL-5.

2.3. Procedure

Data was collected through an online survey, with recruitment coordinated by three psychological research centres and outpatient clinics, enrolling participants between 2022 and 2024. Informed consent was obtained from all participants. Those who refused to provide their informed consent were not permitted to participate. Overall, participants provided

Table 1. Demographic Characteristics.

	German subsample $n = 283$		Arabic subsample $n = 295$		Total sample $n = 578$	
Characteristic	*n* (%)	*M (SD)*	*n* (%)	*M (SD)*	*n* (%)	*M (SD)*
Age[a]		35.33 (12.36)		34.41 (10.90)		34.86 (11.64)
Gender						
Male	92 (32.51)		183 (62.03)		275 (47.58)	
Female	188 (66.43)		111 (37.63)		299 (51.73)	
Prefer not to say	3 (1.06)		1 (0.34)		4 (0.69)	
Country of origin						
Germany	283 (100.00)		1 (0.34)		284 (49.13)	
Middle Eastern countries	–		250 (84.75)		250 (43.25)	
African countries	–		37 (12.54)		37 (6.40)	
Other countries	–		1 (0.34)		1 (0.17)	
Psychological disorder/ distress[b]						
No	230 (81.27)		219 (74.24)		449 (77.68)	
Yes	53 (18.73)		76 (25.76)		129 (22.32)	
Psychological or psychiatric treatment						
No	244 (86.22)		242 (82.03)		486 (84.08)	
Yes	39 (13.78)		53 (17.97)		92 (15.92)	
PTSD[c]						
Sum score	283 (100.00)	14.13 (15.72)	282 (95.59)	26.63 (19.76)	565 (97.75)	20.37 (18.90)
Traumatic events[d]						
At least one	233 (82.33)		261 (88.47)		494 (85.47)	
Total per person		2.89 (2.66)		5.73 (4.16)		4.34 (3.78)

Note. Percentages are based on the total (sub)sample. Totals may vary slightly from 100% due to rounding and missing data. [a]in years; [b]participants answering *yes* to psychological disorder/ distress were classified as clinical; *no* as non-clinical; [c]according to the PCL-5; [d]according to the LEC-5.

sociodemographic data and completed the LEC-5 and PCL-5, and they were compensated for their participation. The study was approved by the local ethics committee (ethics committee number: 2123) in October 2021.

2.4. Data analysis

All statistical analyses were conducted using R 4.1.0 (R Core Team, 2021). For the CFAs and MGCFAs, the lavaan package (Rosseel, 2012) was used. Due to the current limitations in handling ordinal variables with Full Information Maximum Likelihood (FIML; Enders & Bandalos, 2001) in lavaan, listwise deletion was applied to cases with missing PCL-5 data. This conservative approach was chosen to ensure valid estimations under the given model assumptions.

First, we computed descriptive statistics for the sociodemographic, psychological distress, and current treatment data, focusing especially on the descriptive analysis of stressful or traumatic life events assessed using the LEC-5, and PTSD symptom severity measured with the PCL-5. Then, the internal consistency of the PCL-5 was assessed using Cronbach's alpha to ensure an adequate level of reliability for subsequent structural modelling.

To evaluate the structural validity of the PCL-5 in a cross-cultural comparison, we conducted separate CFAs for both, the German and Arabic subsamples, examining the DSM-5 model, Dysphoria model, Dysphoric arousal model, Anhedonia model, Externalising behaviour model, and Hybrid model. The primary aim of these separate CFAs was to examine whether the same models provided the best fit within each subsample, thus offering preliminary insights into the cross-linguistic and cross-cultural comparability of the PCL-5's structural validity. Subsequently, the same set of models was evaluated in a CFA conducted on the total sample. The best-fitting model from this analysis was used as the baseline model for the subsequent MGCFA. Based on the recommendations by Somaraju et al. (2022) measurement invariance of the PCL-5 across the German and Arabic subsamples was examined using a hierarchical procedure. Specifically, we sequentially tested for configural invariance(equal factor structures across the subsamples), threshold invariance (given the ordinal nature of the data, thresholds were also tested for equality across subsamples), metric invariance (equal factor loadings across the subsamples), scalar invariance (equal loadings, thresholds, and intercepts across the subsamples), and full invariance (equal loadings, thresholds, intercepts, and residuals across the subsamples) (Somaraju et al., 2022). A comprehensive overview on the interpretation of measurement invariance levels in cross-cultural research can be found in Milfont and Fischer (2010). Due to the ordinal nature of the PCL-5 variables, CFA and MGCFA were conducted using a Weighted least square mean and variance (WLSMV) adjusted estimator with theta parameterisation, as recommended by Muthén and Christoffersson (1981), to effectively handle categorial data and address data skewness. Since well-fitting models may sometimes yield significant χ^2 values when the factor structure is complex (Bentler & Bonett, 1980), we further considered the recommendations of Hu and Bentler (1999) for assessing model fit. Models in CFA and MGCFA are

considered well-fitting if the comparative fit index (*CFI*) and Tucker–Lewis index (*TLI*) are ≥ .95, the root mean square error of approximation (*RMSEA*) is < .06, and the standardised root mean square residual (*SRMR*) is ≤ .08 (Hu & Bentler, 1999). In the model fit comparison, differences were considered negligible when the fit indices were as follows: *ΔCFI* < .002, *ΔTLI* < .01, *ΔRMSEA* < .015, and *ΔSRMR* < .030 (Chen, 2008; Hu & Bentler, 1999; Meade et al., 2008). To compare nested models in MGCFA, we used the Satorra–Bentler test (Satorra & Bentler, 2001), which is robust against violations of normal distribution.

3. Results

3.1. Descriptive statistics

The participants in the present study were either German- ($n = 295$) or Arabic-speaking ($n = 283$), and all resided in Germany to control for institutional conditions. Of the Arabic-speaking subsample, $n = 121$ participants indicated that they were residing in Germany as refugees or asylum seekers, while $n = 68$ reported having immigrated for professional or personal reasons. A considerable proportion of the Arabic participants were from Syria, while the remaining participants came from various African and Middle Eastern countries (see Table 1 for details). The German participants were exclusively from Germany. While the total sample demonstrated a balanced gender distribution, the Arabic and German subsamples exhibited significant differences ($\chi^2(1) = 48.46$, $p < .001$, for details see Table 1). In total, 22.32% of participants reported suffering from a mental disorder or psychological distress, with no significant difference observed between the two subsamples ($\chi^2(1) = 3.73$, $p = .054$). Similarly, a slightly lower proportion of the total sample (15.92%) reported receiving psychotherapeutic and/or psychiatric treatment, also with no meaningful difference between the two subsamples ($\chi^2(1) = 1.59$, $p = .207$). All participants reported experiencing at least one stressful life event. Among them, 85.47% of the total sample (88.47% of Arabic participants and 82.33% of German participants) reported exposure to at least one potentially traumatic event, as assessed with the LEC-5. The remaining reported events were reviewed for plausibility and included experiences such as bullying, parental separation, or the loss of a pet. Furthermore, based on the LEC-5, the total sample reported an average of $M = 4.34$ ($SD = 3.78$) traumatic events experienced or witnessed per person. Specifically, the Arabic subsample reported an average of $M = 5.73$ ($SD = 4.16$) traumatic events per person, approximately twice as many as the German subsample ($M = 2.89$, $SD = 2.66$; $t(502.66) = 9.80$, $p < .001$; see Table 1). In terms of trauma type, the most commonly reported events included transport accidents and severe human suffering. However, the distribution of trauma types varied notably between the German and Arabic subsamples. In the Arabic subsample, the most commonly reported events were man-made, such as a fire or explosion, combat or exposure to a war zone. By contrast, in the German subsample, the most common events were accidental, such as transport accidents or life-threatening illness or injury (see SM2). Regarding the PTSD symptom severity, the total sample (n = 565) showed a mean PCL-5 score of $M = 20.37$ ($SD = 18.90$; range = [0, 80]). Mean PCL-5 scores were significantly higher in the Arabic subsample ($M = 26.63$, $SD = 19.76$, range = [0; 80]) compared to the German subsample ($M = 14.13$, $SD = 15.72$, range = [0; 71]; $W = 55{,}054$, $p < .001$), indicating a greater PTSD symptom burden in the Arabic subsample.

3.2. Internal consistency

The internal consistency of the PCL-5 demonstrated excellent reliability in the total sample (Cronbach's $\alpha = .96$, 95% CI [.95–.96]), as well as in the German (Cronbach's $\alpha = .95$, 95% CI [.94–.96]) and Arabic (Cronbach's $\alpha = .96$, 95% CI [.95–.96]) subsamples. The average inter-item correlation was $M(r) = 0.54$, with negligible differences between the German and Arabic versions of the PCL-5. Notably, inter-factor correlations allowed us to measure PTSD as a cohesive construct.

3.3. Structural validity

To assess structural validity, the most common six models (see SM1) were tested using separate and overall CFAs. The fit indices for the separate CFAs of the German and Arabic versions of the PCL-5 are presented in Table 2, and the fit indices for the overall CFA are shown in Table 3. Examining the German version of the PCL-5, the CFAs indicated an excellent model fit for both the Hybrid and Anhedonia models, while all models were considered to have an adequate fit. Differences in fit indices comparing the Anhedonia and Hybrid models indicated a negligible change in model fit that marginally favoured the Anhedonia model over the Hybrid model.

For the Arabic version of the PCL-5, the CFA results indicated an acceptable model fit for all models tested, but the Hybrid model was favoured. Given the generally good model fit, we investigated the changes in fit indices, but we found that the *ΔCFI*, *ΔTLI*, *ΔSRMR*, and *ΔRMSEA* showed negligible changes in model fit. The CFA of the total sample confirmed the findings of a good to excellent model fit for both the Anhedonia and Hybrid models, with a marginal preference for the Hybrid over the Anhedonia model.

Table 2. Overview of Model Fit Indices from the CFAs per Subsample.

German subsample													
Model	*n*	*npar*	χ^2	*df*	*p*	*CFI*	*ΔCFI*	*SRMR*	*ΔSRMR*	*TLI*	*ΔTLI*	*RMSEA* [90% CI]	*ΔRMSEA*
DSM-5	283	106	266.375	164	<.001	.975	–	.055	–	.972	–	.071 [.062–.080]	–
Dysphoria	283	106	263.386	164	<.001	.975	–	.057	–	.971	–	.071 [.062–.080]	–
Externalising behaviour	283	115	202.608	155	.006	.981	–	.050	–	.976	–	.065 [.056–.074]	–
Dysphoric arousal	283	110	205.452	160	.009	.981	–	.050	–	.978	–	.063 [.053–.072]	–
Hybrid	283	121	127.076	149	.903	.989	–	.040	–	.987	–	.049 [.038–.059]	–
Anhedonia	283	115	131.364	155	.916	.990	.000	.041	−.001	.988	−.001	.047 [.036–.057]	.002
Arabic subsample													
Model	*n*	*npar*	χ^2	*df*	*p*	*CFI*	*ΔCFI*	*SRMR*	*ΔSRMR*	*TLI*	*ΔTLI*	*RMSEA* [90% CI]	*ΔRMSEA*
Dysphoria	282	106	286.397	164	<.001	.970	–	.047	-	.965	–	.085 [.077–.094]	–
Dysphoric arousal	282	110	204.159	160	.010	.979	–	.042	-	.975	–	.072 [.063–.081]	–
DSM-5	282	106	208.335	164	.011	.979	–	.042	-	.976	–	.071 [.062–.080]	–
Externalising behaviour	282	115	185.597	155	.047	.980	–	.040	-	.976	–	.071 [.062–.080]	–
Anhedonia	282	115	156.335	155	.455	.985	–	.037	-	.981	–	.063 [.053–.072]	–
Hybrid	282	121	135.745	149	.774	.986	−.002	.034	.003	.983	−.001	.060 [.050–.070]	.022

Note. Models are presented in hierarchical order, with the best-fitting model at the bottom. ΔCFI and ΔRMSEA represent differences in fit indices relative to the immediately preceding model with a slightly poorer fit. *n:* number of participants; *npar:* number of parameters; χ^2: chi-square; *df*: degree of freedom; *CFI*: comparative fit index; *SRMR*: standardised root mean square residual; *TLI*: Tucker–Lewis index; *RMSEA*: root mean square error of approximation.

In summary, the separate and the total sample CFAs suggest that both the Anhedonia and Hybrid models have a good to excellent model fit. Consequently, we tested measurement invariance by comparing the German and Arabic versions of the PCL-5 based on both the Hybrid and Anhedonia models via MGCFA, which involved a series of hierarchical model tests.

3.4. Measurement invariance

We first assessed configural and threshold invariance based on the Anhedonia model, then we tested metric, scalar, and full invariance. The Anhedonia model confirmed configural and threshold invariance, with fit indices meeting the cut-off criteria. Notably, in the MGCFA of the Anhedonia model, metric invariance was evaluated as given. Testing scalar invariance, $CFI = .986$, $TLI = .986$, $SRMR = .044$ and $RMSEA = .053$, 90% CI [.046–.059] (see Table 4), suggested a reasonably good model fit, $\chi^2(df = 378) = 519.095$, $p < .001$. Additionally, the comparison of the metric and scalar models, indicated by $\Delta\chi^2(df = 14) = 91.888$, $p < .001$, led to the rejection of the assumption of scalar and full invariance.

Following the same procedure mentioned above, the results of the MGCFA based on the Hybrid model confirmed both configural and threshold invariance (see Table 4 for fit statistics). For metric invariance, the model test indicated a good model fit for the metric model, while the remaining fit indices suggested a good model fit for the scalar model: $CFI = .986$, $TLI = .985$, $SRMR = .043$, $RMSEA = .054$, 90% CI [.047–.060]. However, scalar invariance was rejected in the MGCFA based on the Hybrid model, as indicated by $\chi^2(df = 364) = 490.434$, $p < .001$ and $\Delta\chi^2(df = 13) = 91.135$, $p < .001$, when comparing the metric to the scalar invariance model. Full invariance was similarly rejected due to the rejection of scalar invariance in the MGCFA based on the Anhedonia and Hybrid models.

In the MGCFA, the German and Arabic versions of the PCL-5 did not show scalar invariance. The fit indices (*CFI* and *TLI*) also increased up to metric invariance and then decreased, while *RMSEA* reached its lowest value during metric invariance testing and increased thereafter (see Table 4). This suggested a good model fit for the configural and threshold invariance model, assuming that metric invariance did not degrade the model fit. Testing the scalar invariance model revealed significant chi-square values ($p < .001$) and a significant decrease in model fit ($\Delta\chi^2$) from the metric to the scalar invariance model. This confirmed the rejection of scalar invariance, indicating differences in the intercepts between the two versions of the PCL-5 across subsamples.

4. Discussion

The present study investigated the cross-cultural validation of the German and Arabic versions of the PCL-5 to ensure its cross-cultural comparability.

Initially, the present findings indicate that, while no significant difference was observed between the Arabic and German subsamples in terms of exposure to at least one traumatic event, the two subsamples differ in terms of trauma frequency and trauma type. The Arabic participants reported nearly twice as many traumatic events per person, with a significantly higher rate of man-made trauma. This finding aligns with previous research (Blackmore et al., 2020; Nesterko et al., 2020) that identifies cumulative exposure to interpersonal trauma as a significant risk factor for the development of severe PTSD. Correspondingly, the Arabic subsample displayed higher PCL-5 scores than the German subsample, that reported fewer, predominantly accidental traumatic events.

Table 3. Overview of Model Fit Indices from the CFAs in the Total Sample.

	Total sample												
Model	*n*	*npar*	χ^2	*df*	*p*	*CFI*	*ΔCFI*	*SRMR*	*ΔSRMR*	*TLI*	*ΔTLI*	*RMSEA* [90% CI]	*ΔRMSEA*
Dysphoria	565	106	434.458	164	<.001	.974	–	.042	–	.970	–	.081 [.075–.087]	–
DSM-5	565	106	349.954	164	<.001	.980	–	.038	–	.976	–	.072 [.066–.078]	–
Dysphoric arousal	565	110	322.838	160	<.001	.981	–	.037	–	.977	–	.071 [.065–.077]	–
Externalising behaviour	565	115	314.054	155	<.001	.980	–	.037	–	.976	–	.073 [.067–.079]	–
Anhedonia	565	115	193.955	155	.018	.989	–	.030	–	.986	–	.055 [.048–.061]	–
Hybrid	565	121	178.296	149	.051	.989	.000	.029	.001	.986	.000	.055 [.048–.061]	.000

Note. Models are presented in hierarchical order, with the best-fitting model at the bottom. ΔCFI and ΔRMSEA represent differences in fit indices relative to the immediately preceding model with a slightly poorer fit. *n:* number of participants; *npar:* number of parameters; χ^2: chi-square; *df*: degree of freedom; *CFI*: comparative fit index; *SRMR*: standardised root mean square residual; *TLI*: Tucker–Lewis index; *RMSEA*: root mean square error of approximation.

These findings underscore the importance of a robust, culturally sensitive validation of the PCL-5 to enable a valid interpretation of (the observed) differences in symptom severity and prevalence.

The internal consistency of the PCL-5 was found to be excellent, supporting its status as a psychometrically robust instrument (Blevins et al., 2015; Brahim et al., 2022; Krüger-Gottschalk et al., 2017) and justifying further investigation of its structural validity.

The subsample CFAs and the total sample CFA demonstrated an adequate fit for all tested models. The Hybrid and Anhedonia models exhibited the best model fit, consistent with the conceptual overlap between these six to seven factor models. Both models include the core symptom clusters of PTSD such as *intrusion, avoidance, negative affect, anhedonia*, as well as *dysphoric* and *anxious arousal*. The Hybrid model further differentiates *dysphoric arousal* by isolating *externalising symptoms* (e.g. irritability, risk-taking behaviour), thereby recognising that heightened negative affect or cognitive distress does not necessarily lead directly to behavioural dysregulation.

While the Anhedonia and Hybrid models are frequently identified as the most suitable structures in both the Arabic and German PCL-5 (e.g. Ibrahim et al., 2019), this alone does not justify conclusions about cross-cultural comparability (Brown et al., 2015). Moreover, an examination of the PCL-5's measurement invariance is essential to ensure the comparability of outcomes in its global application. In light of the ongoing political unrest and global refugee movements, the evaluation of the cross-cultural applicability of the PCL-5, particularly between European and Middle Eastern populations, is highly relevant. Addressing this, the present study applied MGCFA to test measurement invariance across the German and Arabic versions of the PCL-5. The results of the MGCFA, based on the Anhedonia and Hybrid models, supported configural, metric, and threshold invariance, underscoring the instrument's conceptual robustness in this cross-linguistic and potentially cross-cultural contexts. The confirmed metric invariance suggests that the PCL-5 items exhibit a comparable correlation with underlying PTSD factors in both the German and Arabic versions. However, the lack of scalar invariance suggests that some items may function differently across both, the Arabic and the German versions. In particular, differences in item intercepts have the potential to lead to different PCL-5 sum scores, even when the underlying level of PTSD severity is equivalent across groups. The underlying causes of the differences in item intercepts cannot be definitively ascertained based on the present findings. One possible explanation for this phenomenon is that lacking scalar invariance reflects variations in the meaning-making and expression of PTSD symptoms across the Arabic and the German languages and cultures (Hosny et al., 2023). For instance, certain symptoms of PTSD may carry different connotations, are more or less socially acknowledged, or are interpreted differently in terms of severity or salience, potentially influencing the item responses (Hinton & Lewis-Fernández, 2011; Hosny et al., 2023). Therefore, a comparison of the PCL-5

Table 4. Fit Statistics for the MGCFA of the German and Arabic versions of the PCL-5 across German and Arabic Subsamples

Anhedonia model	*n*	*npar*	χ^2	*df*	*p*	*CFI*	*SRMR*	*TLI*	*RMSEA* [90% *CI*]
Configural model	565	230	287.700	310	.814	.987	.039	.985	.055 [.047–.062]
Threshold model	565	190	346.418	350	.544	.987	.039	.986	.051 [.045–.058]
Metric model	565	176	427.206	364	.012	.989	.044	.989	.047 [.040–.054]
Scalar model	565	162	519.095	378	<.001	.986	.044	.986	.053 [.046–.059]
Full model	565	142	519.095	398	<.001	.985	.044	.986	.052 [.046–.058]
Hybrid model	***n***	***npar***	χ^2	***df***	***p***	***CFI***	***SRMR***	***TLI***	***RMSEA*** [90% *CI*]
Configural model	565	242	262.820	298	.930	.988	.037	.985	.054 [.047–.061]
Threshold model	565	202	321.539	338	.732	.988	.037	.987	.051 [.044–.058]
Metric model	565	189	401.350	351	.033	.989	.043	.988	.048 [.041–.055]
Scalar model	565	176	490.434	364	<.001	.986	.043	.985	.054 [.047–.060]
Full model	565	156	490.434	384	<.001	.985	.043	.985	.053 [.047–.059]

Note. *n:* number of participants; *npar:* number of parameters; χ^2: chi-square; *df:* degree of freedom; *CFI*: comparative fit index; *SRMR*: standardised root mean square residual; *TLI*: Tucker–Lewis index; *RMSEA*: root mean square error of approximation.

sum scores of the German and Arab versions may result in an inaccurate interpretation, as the observed differences may reflect a measurement bias rather than actual differences in symptom severity. A frequently applied approach in current research involves the utilisation of sample-specific cut-off adjustments (Pettrich et al., 2025), derived from sensitivity and specificity analyses. Although such adjustments have been implemented empirically in previous studies, measurement invariance of the PCL-5 German and Arabic version had not yet been formally examined. The present findings are consistent with these prior adaptations and offer preliminary empirical evidence that validates their efficacy.

4.1. Limitations

Due to the specificity of the examined sample, particularly the Arabic subsample, the scalar non-invariance observed in this study suggests that PCL-5 outcomes may not be fully comparable between German individuals and Arabic individuals with refugee or migration backgrounds. Although the total sample was recruited in Germany to align living conditions and ensure system standardisation, the Arabic subsample differs from the German subsample in terms of sociocultural homogeneity. Specifically, the Arabic subsample primarily includes migrants and refugees from various countries of origin, resulting in increased heterogeneity within the subsample and potentially contributing to selection bias. Accordingly, differences between the subsamples extend beyond mere linguistic or cultural variation. Due to this sociocultural heterogeneity and other confounding factors, the results must be interpreted with caution, as Arabic culture should not be treated as a homogenous construct. Thus, no conclusions should be drawn regarding the comparability of PCL-5 outcomes across all Arabic-speaking populations. Additionally, processes of cultural adaptation in the host country may have influenced symptom reporting, potentially leading to an underestimation of measurement variance in the cross-cultural comparison and overly optimistic conclusions.

A selection bias is likely introduced by conducting the (MG)CFA as a complete-case analysis by including only participants with complete PCL-5 data. This may limit representativeness of the sample, as missing data, which is not entirely random, may systematically exclude certain subgroups. Although alternative methods for handling missing data, such as FIML, are well-established in structural equation modelling, FIML is currently not fully supported for ordinal (categorical) indicators in lavaan (Rosseel, 2012). Therefore, we opted for a conservative approach using WLSMV estimation in lavaan to ensure appropriate modelling of ordinal data. This approach also minimises potential bias from imputation and enhances internal validity, but at the cost of a reduced sample size and a potential increase in selection bias if data is not missing completely at random. Consequently, this limits ecological validity and the generalizability of our findings.

The lack of scalar invariance often prompts testing for partial scalar invariance, which involves freeing intercepts for items exhibiting non-invariance (Putnick & Bornstein, 2016). However, as discussed above, the limited comparability of the subsamples can complicate such analyses. Additionally, when assessing partial scalar invariance, it is assumed that only a limited number of items vary across groups. Conversely, the absence of scalar invariance alongside robust metric invariance demonstrated in the present study suggests that a considerable number of items likely exhibit non-invariance. Therefore, the examination of partial invariance appears to be unwarranted in this context.

Another limitation of the present study is that it considered only those factor models that had already been assessed in both the German and Arabic versions of the PCL-5. While this approach ensures a methodologically well-founded model selection, it does not claim to be exhaustive. For instance, the parsimonious and psychometrically robust two-factor bifactor model proposed and tested by Pettrich et al. (2024) in the German PCL-5 version was not included in the analyses.

4.2. Clinical implications

Parsimonious models, such as the DSM-5 model, have demonstrated clinical advantages (e.g. Pettrich et al., 2024). The reduced number of factors in general results in an increased number of items per factor, thereby enhancing model stability for parsimonious models. Thus two- to four-factor models have been proposed to enhance clinical utility (e.g. Schmitt et al., 2018). Conversely, more complex models comprising numerous narrowly defined factors are challenging to implement in clinical practice, particularly when individual factors are represented by fewer than three items (e.g. avoidance or anxious arousal). This has been demonstrated to be associated with diminished psychometric stability and limited interpretability (Rasmussen et al., 2019). Nonetheless, while complex factor models may complicate clinical implementation, they allow for a more nuanced representation of PTSD symptomatology. For instance, the Hybrid and Anhedonia models differentiate symptom clusters (e.g. intrusions, avoidance, negative mood/cognition, hyperarousal), supporting more precise diagnosis. Consistently, these two multidimensional models demonstrated superior model fit compared to parsimonious factor models, such as the DSM-5 model, in the present study.

As previously indicated, the lack of scalar invariance in the superior models examined in MGCFA indicates that PCL-5 cut-off scores may necessitate sample-specific adaptation. Current research recommended cut-off scores ranging from 22 to 49. For instance, Krüger-Gottschalk et al. (2017) proposed a cut-off of 31–33 for German samples, while Ibrahim et al. (2018) advocated a cut-off of 23 for Arabic samples from (post-)conflict regions. The adaptation of cut-off scores to the cultural, linguistic, and contextual characteristics of the target population, as described previously, has the potential to enhance the accuracy and comparability of PTSD assessments across groups (see also Ibrahim et al., 2018). In heterogeneous populations, sample-specific cut-off adaptations remain the most pragmatic approach to minimise diagnostic bias, ensuring cultural sensitivity and maintaining the comparability to ensure the global validity of the PCL-5.

4.3. Future research

Future research should aim to identify models that achieve an optimal balance between good model fit, model parsimony, clinical utility, and diagnostic precision.

In order to consider clinical implications, it is essential that future research include parsimonious models in their measurement invariance analysis, considering all recent factor models for PCL-5.

The present findings are currently limited to German- and Arabic-speaking populations in Europe. Future research should include more diverse populations (e.g. Hansen et al., 2023) to enhance generalizability and validate diagnostic instruments that account for population-specific differences in symptom perception and expression, enabling accurate cross-group comparisons (Kleim & Ehlers, 2008; Wortmann et al., 2016). To ensure diagnostic accuracy across diverse populations and contexts, measurement invariance analyses of the PCL-5 should be extended. This will improve prevalence estimates, facilitate the identification of treatment needs, and enhance access to psychotherapeutic care.

While the present study primarily focused on measurement invariance between the German and Arabic versions, other studies have examined either the factor structure or the convergent and divergent validities of the PCL-5 separately within the German and Arabic samples (e.g. Ibrahim et al., 2018; Krüger-Gottschalk et al., 2017). Thus the present conclusions are based on a combination of the current research, and the present study findings. Due to the methodological scope of the analyses, a full integration of these additional evaluations was not feasible within the present study. Therefore, to strengthen the basis from speculative to concrete recommendations on cut-off adaptations, we intend to expand the present study of measurement invariance by conducting a comprehensive psychometric evaluation of the PCL-5, encompassing assessments of construct validity, convergent and discriminant validity, as well as criterion validity. In this regard, sensitivity, specificity, and ROC analyses will be employed to ascertain subgroup-specific cutoff scores for the PCL-5 German and Arabic version.

To comprehensively assess measurement invariance in cross-cultural PTSD assessment, future research should extend such designs beyond a single instrument, such as the German and Arabic versions of the PCL-5. While preliminary evidence supports the hypothesis of configural, metric, and scalar invariance of the Impact of Event Scale–Revised (IES-R) across gender, age, and marital status (e.g. Ali et al., 2022), cross-linguistic or cross-cultural invariance for IES-R and also for clinical interviews like the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5; Weathers et al., 2013b), between English or other Western language and Arabic versions, remains largely unexplored. Expanding measurement invariance testing to a broader range of tools would therefore enhance understanding of the cross-cultural applicability of the Western conceptualisation of PTSD.

4.4. Conclusion

Overall, both the German and Arabic versions of the PCL-5 are reliable diagnostic tools for assessing PTSD symptoms, with a six-to-seven-factor structure (Anhedonia and Hybrid models) being confirmed for both versions by single-group CFAs. Moreover, the present study provides important evidence for metric measurement invariance of the PCL-5 across the German and Arabic populations, allowing meaningful comparisons of PTSD symptom structure between these populations. However, the scalar non-invariance suggests that the sum scores are not directly comparable, as it is debatable whether the strength in symptom experience and expression of PTSD in PCL-5 differ between German and Arabic. Addressing this, future research is needed to validate context-sensitive, population-adapted cut-off scores to improve diagnostic accuracy. In addition, comprehensive psychometric research is needed, including the investigation of measurement invariance in other PTSD instruments in several populations for a more global view on PTSD assessment adequacy.

In a broader context, our findings underscore the critical need for cultural, linguistic, and contextual sensitivity in the assessment of PTSD to avoid bias

in epidemiologic research (e.g. prevalence estimates) and clinical practice (e.g. assessment of mental health care needs and identification of appropriate treatment), particularly when applying Western-developed instruments to populations with divergent cultural backgrounds. Ultimately, adopting flexible, context-adapted approaches will enhance both psychometric robustness and clinical utility, ensuring fair, valid cross-cultural PTSD diagnoses and better-informed treatment decisions.

Open scholarship

This article has earned the Center for Open Science badge for Open Materials. The materials are openly accessible at https://github.com/CharinaCLueder/ccv_pcl5.

Acknowledgments

We would like to express our gratitude to the students who assisted in data collection as part of their thesis projects. We also extend our sincere thanks to I. Rickert, representing the Center for Integrative Psychiatry (ZIP) in Lübeck, for their support with data collection. CCL: Research question, conceptualisation, project administration and implementation, data acquisition, resource acquisition, methodology, formal analysis, data curation, data integration, writing (original draft, final draft), visualisation. ME: Resource acquisition, supervision, writing (review and editing). NB: Supervision, methodology, writing (review and editing). AS: Data acquisition, data curation, data integration, methodology. HG: Data acquisition, data curation, resource acquisition. YN: Data acquisition, data curation, resource acquisition. TM: Resource acquisition, supervision, writing (review and editing).

Disclosure statement

No potential conflict of interest was reported by the authors.

Open practices statement

Data supporting the results of this study are available upon request. Requests for access to the data can be directed to the corresponding author at charina.lueder@uni-saarland.de. The R scripts used for the analyses reported in the present study are publicly available at https://github.com/CharinaCLueder/ccv_pcl5.

Clinical trial registration

The study reported in this article was not formally preregistered.

Patient consent statement

Informed consent was obtained from all participants prior to their inclusion in the study. The participants' informed consent form, English version, is provided in SM3.

Permission to reproduce material from other sources

The German and Arabic versions of the PCL-5 were either publicly accessible or procured directly from the original authors (who also served as co-authors of this study) and have been cited in accordance with APA standards.

Ethical standards statement

The Ethics Committee of the Faculty of Empirical Human Sciences and Economics at Saarland University approved the ethics application for this study in October 2021 (Ethics Committee reference number: 2123). All procedures were conducted in accordance with the ethical standards outlined in the Declaration of Helsinki. Prior to participation, all individuals were required to provide written informed consent (see SM3 for English version).

ORCID

Charina C. Lüder http://orcid.org/0009-0005-0422-046X
Monika Equit http://orcid.org/0000-0002-7213-5962
Antje Schönfelder http://orcid.org/0000-0002-2009-0512
Heide Glaesmer http://orcid.org/0000-0002-3752-7653
Yuriy Nesterko http://orcid.org/0000-0002-8296-0958
Tanja Michael http://orcid.org/0000-0002-2409-3817

References

Ali, A. M., Al-Amer, R., Kunugi, H., Stanculescu, E., Taha, S. M., Saleh, M. Y., Alkhamees, A. A., & Hendawy, A. O. (2022). The Arabic version of the Impact of Event Scale-Revised: Psychometric evaluation among psychiatric patients and the general public within the context of COVID-19 outbreak and quarantine as collective traumatic events. *Journal of Personalized Medicine*, *12*(5), 681. https://doi.org/10.3390/jpm12050681

AlShawi, A. F. (2018). Prevalence of posttraumatic stress disorders among a sample of internally displaced persons in Iraq: A preliminary study. *Journal of Community Medicine & Health Education*, *8*(2), Article 1000599. https://doi.org/10.4172/2161-0711.1000599

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

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. https://doi.org/10.1016/j.jpsychires.2014.10.012

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Blackmore, R., Boyle, J. A., Fazel, M., Ranasinha, S., Gray, K. M., Fitzgerald, G., Misso, M., & Gibson-Helm, M. (2020). The prevalence of mental illness in refugees and asylum seekers: A systematic review and meta-analysis. *PLOS Medicine*, *17*(9), e1003337. https://doi.org/10.1371/journal.pmed.1003337

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*(6), 489–498. https://doi.org/10.1002/jts.22059

Brahim, C. B., El Kefi, H., Awissawi, A., Kabtni, W., Baatout, A., Bouizouita, I., Bouguerra, C., Eddif, S., & Oumaya, A. (2022). Validation en arabe littéraire chez les militaires tunisiens de la Post-Traumatic Stress Disorder Checklist for DSM-5. *Pan African Medical Journal*, *43*(28), Article 28. https://doi.org/10.11604/pamj.2022.43.28.35869

Brown, G. T. L., Harris, L. R., O'Quin, C. R., & Lane, K. (2015). Using multigroup confirmatory factor analysis to evaluate cross-cultural research: Identifying and understanding non-invariance. *International Journal of Research and Method in Education*, *40*(1), 66–90. https://doi.org/10.1080/1743727X.2015.1070823

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in Cross-National Research. *Annual Review of Sociology*, *40*(1), 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Elhai, J. D., Biehn, T. L., Armour, C., Klopper, J. J., Frueh, B. C., & Palmieri, P. A. (2011). Evidence for a unique PTSD construct represented by PTSD's D1–D3 symptoms. *Journal of Anxiety Disorders*, *25*(3), 340–345. https://doi.org/10.1016/j.janxdis.2010.10.007

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. doi:10.1207/S15328007SEM0803_5

Fazel, M., Wheeler, J., & Danesh, J. (2005). Prevalence of serious mental disorder in 7000 refugees resettled in Western countries: A systematic review. *The Lancet*, *365*(9467), 1309–1314. https://doi.org/10.1016/S0140-6736(05)61027-6

Forkus, S. R., Raudales, A. M., Rafiuddin, H. S., Weiss, N. H., Messman, B. A., & Contractor, A. A. (2022). The Posttraumatic Stress Disorder (PTSD) Checklist for DSM–5: A systematic review of existing psychometric evidence. *Clinical Psychology: Science and Practice*, *30*(1), 110–121. https://doi.org/10.1037/cps0000111

Georgiadou, E., Morawa, E., & Erim, Y. (2017). High manifestations of mental distress in Arabic asylum seekers accommodated in collective centers for refugees in Germany. *International Journal of Environmental Research and Public Health*, *14*(6), 612. https://doi.org/10.3390/ijerph14060612

Hansen, M., Vaegter, H. B., Ravn, S. L., & Andersen, T. E. (2023). Validation of the Danish PTSD Checklist for DSM-5 in trauma-exposed chronic pain patients using the Clinician-Administered PTSD Scale for DSM-5. *European Journal of Psychotraumatology*, *14*(1), Article 2179801. https://doi.org/10.1080/20008066.2023.2179801

Hinton, D. E., & Lewis-Fernández, R. (2011). The cross-cultural validity of posttraumatic stress disorder: implications for DSM-5. *Depression and anxiety*, *28*(9), 783–801. doi:10.1002/da.20753

Hoffman, J., Ben-Zion, Z., Arevalo, A., Duek, O., Greene, T., Hall, B. J., Harpaz-Rotem, I., Liddell, B., Locher, C., Morina, N., Nickerson, A., Pfaltz, M. C., Schick, M., Schnyder, U., Seedat, S., Shatri, F., Fong Sit, H., von Känel, R., & Spiller, T. R. (2022). Mapping the availability of translated versions of posttraumatic stress disorder screening questionnaires for adults: A scoping review. *European journal of psychotraumatology*, *13*(2), 2143019. doi:10.1080/20008066.2022.2143019

Hosny, N., Bovey, M., Dutray, F., & Heim, E. (2023). How is trauma-related distress experienced and expressed in populations from the Greater Middle East and North Africa? A systematic review of qualitative literature. *SSM - Mental Health*, *4*, 100258. doi:10.1016/j.ssmmh.2023.100258

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), 1–55. https://doi.org/10.1080/10705519909540118

Ibrahim, H., Catani, C., Ismail, A. A., & Neuner, F. (2019). Dimensional structure and cultural invariance of DSM-5 post-traumatic stress disorder among Iraqi and Syrian displaced people. *Frontiers in Psychology*, *10*, Article 1505. https://doi.org/10.3389/fpsyg.2019.01505

Ibrahim, H., Ertl, V., Catani, C., Ismail, A. A., & Neuner, F. (2018). The validity of Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) as screening instrument with Kurdish and Arab displaced populations living in the Kurdistan region of Iraq. *BMC Psychiatry*, *18*(1), 259. https://doi.org/10.1186/s12888-018-1839-z

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Benjet, C., Bromet, E. J., Cardoso, G., Degenhardt, L., de Girolamo, G., Dinolova, R. V., Ferry, F., Florescu, S., Gureje, O., Haro, J. M., Huang, Y., Karam, E. G., Kawakami, N., Lee, S., Lepine, J.-P., Levinson, D., … Koenen, K. C. (2017). Trauma and PTSD in the WHO World Mental Health Surveys. *European Journal of Psychotraumatology*, *8*(sup5), 1353383. https://doi.org/10.1080/20008198.2017.1353383

Kleim, B., & Ehlers, A. (2008). Reduced autobiographical memory specificity predicts depression and posttraumatic stress disorder after recent trauma. *Journal of Consulting and Clinical Psychology*, *76*(2), 231–242. https://doi.org/10.1037/0022-006X.76.2.231

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., Karam, E. G., Ruscio, A. M., Benjet, C., Scott, K., Atwoli, L., Petukhova, M., Lim, C. C. W., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Bunting, B., Ciutan, M., Girolamo, G., … Kessler, R. C. (2017). Posttraumatic stress disorder in the World Mental Health Surveys. *Psychological Medicine*, *47*(13), 2260–2274. https://doi.org/10.1017/S0033291717000708

Krüger-Gottschalk, A., Knaevelsrud, C., Rau, H., Dyer, A., Schäfer, I., Schellong, J., & Ehring, T. (2017). The German version of the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Psychometric properties and diagnostic utility. *BMC Psychiatry*, *17*(1), 379. https://doi.org/10.1186/s12888-017-1541-6

Langeland, W., & Olff, M. (2024). Sex and gender in psychotrauma research. *European journal of psychotraumatology*, *15*(1), 2358702. doi:10.1080/20008066.2024.2358702

Liu, P., Wang, L., Cao, C., Wang, R., Zhang, J., Zhang, B., Wu, Q., Zhang, H., Zhao, Z., Fan, G., & Elhai, D.

(2014). The underlying dimensions of DSM-5 posttraumatic stress disorder symptoms in an epidemiological sample of Chinese earthquake survivors. *Journal of Anxiety Disorders*, *28*(4), 345–351. https://doi.org/10.1016/j.janxdis.2014.03.008

Lüder, C. C., Michael, T., Lass-Hennemann, J., Schanz, C. G., Venhorst, A., Meyer, T., & Equit, M. (2023). Moderate-intensity aerobic exercise training as an adjunct to trauma-focused psychotherapy in traumatized refugees and asylum seekers: Study protocol of a randomized controlled trial. *European Journal of Psychotraumatology*, *14*(2), 2251777. https://doi.org/10.1080/20008066.2023.2251777

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of psychological research*, *3*(1), 111–130. doi:10.21500/20112084.857

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, *46*(4), 407–419. https://doi.org/10.1007/BF02293798

Nesterko, Y., Jäckle, D., Friedrich, M., Holzapfel, L., & Glaesmer, H. (2020). Prevalence of post-traumatic stress disorder, depression and somatisation in recently arrived refugees in Germany: An epidemiological study. *Epidemiology and Psychiatric Sciences*, *29*, e40. https://doi.org/10.1017/S2045796019000325

Pettrich, A., Friedrich, M., Nesterko, Y., & Glaesmer, H. (2024). The German PCL-5: Evaluating structural validity in a large-scale sample of the general German population. *European Journal of Psychotraumatology*, *15*(1), 2317055. https://doi.org/10.1080/20008066.2024.2317055

Pettrich, A., Schellong, J., Dyer, A., Ehring, T., Knaevelsrud, C., Krüger-Gottschalk, A., Nesterko, Y., Schäfer, I., & Glaesmer, H. (2025). Beyond one-cutoff-fits-all: determining cutoff values for the PTSD checklist for DSM-5 (PCL-5). *European Journal of Psychotraumatology*, *16*(1), Article 2514878. https://doi.org/10.1080/20008066.2025.2514878

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Quosh, C., Eloul, L., & Ajlani, R. (2013). Mental health of refugees and displaced persons in Syria and surrounding countries: A systematic review. *Intervention Journal of Mental Health and Psychosocial Support in Conflict Affected Areas*, *11*(3), 276–294. https://doi.org/10.1097/WTF.0000000000000013

R Core Team. (2021). *R: A language and environment for statistical computing* (R version 4.1.0). R Foundation for Statistical Computing. https://www.Rproject.org/.

Rasmussen, A., Verkuilen, J., Jayawickreme, N., Wu, Z., & McCluskey, S. T. (2019). When did posttraumatic stress disorder get so many factors? Confirmatory factor models since DSM–5. *Clinical Psychological Science*, *7*(2), 234–248. https://doi.org/10.1177/2167702618809370

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. https://doi.org/10.1007/BF02296192

Schmitt, T. A., Sass, D. A., Chappelle, W., & Thompson, W. (2018). Selecting the "Best" factor structure and moving measurement validation forward: An illustration. *Journal of Personality Assessment*, *100*(4), 345–362. https://doi.org/10.1080/00223891.2018.1449116

Simms, L. J., Watson, D., & Doebbelling, B. N. (2002). Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. *Journal of Abnormal Psychology*, *111*(4), 637–647. https://doi.org/10.1037/0021-843X.111.4.637

Somaraju, A. V., Nye, C. D., & Olenick, J. (2022). A review of measurement equivalence in organizational research: What's old, what's new, what's next? *Organizational Research Methods*, *25*(4), 741–785. https://doi.org/10.1177/10944281211056524

Tsai, J., Harpaz-Rotem, I., Armour, C., Southwick, S. M., Krystal, J. H., & Pietrzak, R. H. (2014). 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*(5), 546–553. https://doi.org/10.4088/JCP.14m09091

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013a). *The Life Events Checklist for DSM-5 (LEC-5)*. U.S. Department of Veterans Affairs. https://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp.

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013b). *The clinician-administered PTSD Scale for DSM-5 (CAPS-5)*.

Wortmann, J. H., Jordan, A. H., Weathers, F. W., Resick, P. A., Dondanville, K. A., Hall-Clark, B., Foa, E. B., Young-McCaughan, S., Yarvis, J. S., Hembree, E. A., Mintz, J., Peterson, A. L., & Litz, B. T. (2016). Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. *Psychological Assessment*, *28*(11), 1392–1403. https://doi.org/10.1037/pas0000260