Università degli Studi di Padova

Department of General Psychology

Bachelor's course in Psychological Science

Final Dissertation

"Factor Analysis and Measurement Invariance of the PTSD Subscale in the International Trauma Questionnaire "

Supervisor

Prof. Andrea Spoto

Co-Supervisor

Dott.ssa Anna Panzeri

Candidate

Mariia Dosaikina

Student ID number

2054049

Academic Year

2023-2024

Abstract

Assessing the psychological impact of trauma is crucial for understanding and addressing its multifaceted consequences, such as Post-Traumatic Stress Disorder (PTSD) and Complex Post-Traumatic Stress Disorder (CPTSD). The International Trauma Questionnaire (ITQ) is a promising tool designed to evaluate and quantify the effects of traumatic experiences. The ITQ has gained significant attention and implementation in clinical practice and psychotraumatology studies due to its parsimony and comprehensible construct structure grounded in the definition of PTSD and CPTSD as present in the 11th version of the International Classification of Diseases (WHO, 2019). The availability of both the original and translated versions of the ITQ has allowed for further research, which has generally confirmed the validity and reliability of the scale in different populations. In previously published papers, the emphasis was placed on the construct validation of the scale. Establishing construct validity is necessary for administering the tool to different groups and populations, as well as for the general replicability of the research. This paper aims to scrutinize the construct validity of ITQ's PTSD subscale by examining the robustness of its factorial structure and its measurement (factorial) invariance across the first and second waves of the COVID-19 pandemic in Italy. The present work creates continuity with previously conducted research by evaluating the latent factorial structure of the PTSD construct. In addition, it extends the pool of evidence for the ITQ's construct validity by confirming the longitudinal measurement invariance of the PTSD subscale.

Table of Contents

1. Introduction

The International Trauma Questionnaire (ITQ) is a concise self-report assessment tool consisting of 18 items. Developed by Cloître et al. in 2018, it aimed at capturing the essential clinical features of Post-Traumatic Stress Disorder (PTSD) and Complex Post Traumatic Stress Disorder (CPTSD) in both clinical and general populations. PTSD and CPTSD are classified as disorders specifically associated with stress that result from exposure to stressful or traumatic events. PTSD is characterized by re-experiencing the traumatic event or events in the form of intrusive memories or nightmares, avoidance of thoughts and memories of the traumatic event, and a persistent sense of current threat. CPTSD meets all diagnostic criteria for PTSD and comprises three additional clinical features that distinguish it from PTSD: problems in affect regulation, difficulty in sustaining close relationships, and negative beliefs about oneself (WHO, 2019). The content and structure of the ITQ align with the definition and diagnostic criteria for PTSD and CPTSD as outlined in the ICD-11 (WHO, 2019). To comprehensively evaluate PTSD and CPTSD and draw a distinction between the two disorders, the ITQ is divided into two subscales: the PTSD scale and the Disturbances in Self-Organization (DSO) scale. Each subscale encompasses three symptom clusters that correspond to the descriptive diagnostic requirements for both disorders:

Post-Traumatic Stress Disorder (PTSD):

1. Re-experiencing (Re)
2. Avoidance (Av)
3. Sense of threat (Th)

Disturbances in Self-Organization (DSO):

1. Affective dysregulation (AD)
2. Negative self-concept (NSC)
3. Disturbances in relationships (DR)

Within each scale, every symptom cluster is represented by two items. Additionally, the ITQ includes six supplementary items (three for each subscale) specifically devised to measure the level of functional impairment associated with either PTSD or CPTSD which allows for two separate scoring systems: categorical scoring for the differential diagnosis of PTSD and CPTSD, and dimensional scoring for the measurement of symptom severity or the evaluation of the effectiveness of implemented treatment (Cloître et al, 2021). Each item is represented by the 5-point Likert scale ranging from 0 to 4 (0 - “Not at all”, 1 - “A little bit”, 2 - “Moderately”, 3 - “Quite a bit”, and 4 -

“Extremely”). A score of 1 or greater is necessary for the item to be endorsed. PTSD is diagnosed if at least one of two items is endorsed in each symptom cluster of the PTSD subscale and if at least one functional impairment item has a sufficient score of 1 or greater. Meanwhile, the diagnosis of CPTSD requires meeting the same criteria for both PTSD and CPTSD subscales simultaneously. The categorical scoring for CPTSD is coherent with the definition of the disorder in the ICD-11 as comprising all the clinical features of PTSD with the addition of disturbances in self-organization. Dimensional scoring of PTSD and CPTSD is based on summing the scores of all the items of either PTSD or DSO subscale respectively with the exclusion of the functional impairment items. Thus, based on such a distinctive scoring system, ITQ is a valuable tool for elucidating the presence of trauma, distinguishing between clinical pictures of CPTSD and PTSD, and estimating the severity and progression of the disorders.

Psychological phenomena, such as trauma, are often latent, that is the underlying constructs of interest are typically unobservable and cannot be measured directly. The precision and factuality of results about a psychological construct depend not only on its theoretical formulation but also on its measurement's validity, making construct validation a fundamental methodology in psychology and other sciences (Flake et al., 2017). While construct validity encompasses the broader idea of whether a tool measures what it is supposed to measure, structural validity is one of the phases of construct validation that scrutinizes the internal structure of the psychometric test and how well its structure and interrelations of items represent the underlying theoretical construct under investigation.

Despite the widespread use and acknowledged usefulness of the ITQ in research and clinical practice, its factorial structure remains debatable and evidence for its longitudinal measurement invariance is still lacking. Therefore, this paper aims to address the gap in the existing literature by evaluating the structural validity of the ITQ's PTSD subscale. The objective of the study is to test different models of the ITQ's factorial structure and to examine the longitudinal measurement invariance of the PTSD subscale in a sample drawn from the Italian general population.

2. Material and Methods

The current study was based on a sample drawn from the adult population of Italy during the COVID-19 pandemic. The data collection process involved administering the ITQ PTSD subscale online, along with a larger battery of psychometric tests within the framework of the broad international study comprising several countries (Bruno et al., 2020). The first online administration took place during the first wave of the COVID-19 pandemic (February - April 2020), while the second administration was conducted during the second wave (October 2020 - January 2021)(Del Re et al., 2024). In total, during the first wave, ITQ was administered to 1048 participants (COVID wave 1 sample). On the other hand, during the second wave, only 544 individuals from the original sample completed the questionnaire (COVID wave 2 sample). Therefore, the COVID wave 2 sample is a subset of the COVID wave 1 sample, and both samples represent two time points in which ITQ was administered. All subjects underwent the administration of the ITQ PTSD subscale, comprising 9 items, each assessed via a 5-point Likert scale. The decision to exclusively administer the PTSD subscale was motivated by several considerations. Primarily, the main objective of the study, which provided the data for this research, was to employ an array of psychometric assessments to foster multiple lines of future research. Additionally, the inclusion of the PTSD subscale of the ITQ was intended to offer a broad overview of the presence of trauma and to indicate possible differences in psychological responses to the first and the second waves of the COVID pandemic within the population rather than to specifically provide the differential diagnosis between PTSD and CPTSD. Despite these constraints, the current paper seeks to provide continuity with the previously conducted validation studies which generally concerned both ITQ subscales. Analysis of the structural validity of the PTSD subscale was performed using R software available to the public domain (R Core Team, 2021).

3. Factorial Structure: Confirmatory Factor Analysis (CFA)

Factorial analysis of a measurement tool consists of two consecutive steps: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). CFA and EFA are members of the broad family of statistical techniques - structural equation models (SEMs). Structural equation modeling allows for testing different hypotheses about a set of measured variables via constructing and testing multivariate models and providing evidence for the interrelationships among the variables (Weston, & Gore, 2006). In SEM, a measurement model is a linear regression model in which the main predictor, the factor, is latent. For a single subject, the simple linear regression equation is defined as follows:

$$y = b_0 + b_1 x + \epsilon$$

where b_0 is the intercept, b_1 is the coefficient, and x is an observed predictor. Similarly, for a single item, the measurement model equation is the following:

$$y = \tau + \lambda\eta + \epsilon$$

where τ is the intercept of the item, λ is the loading of the item, η is an unobserved predictor, and ϵ is the residual of the item.

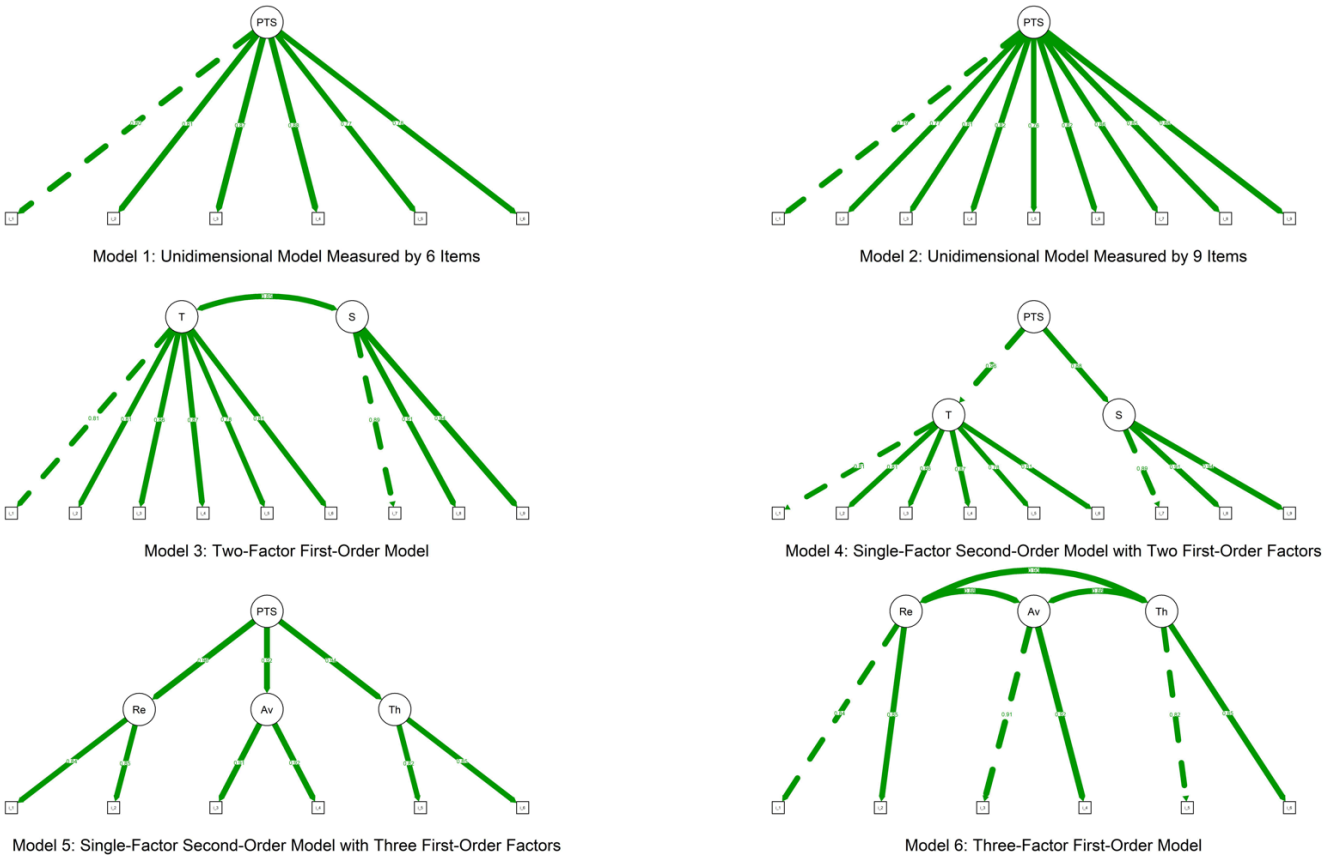
The exploratory factor analysis (EFA) is generally used in the absence of sufficient theoretical or empirical information to hypothesize the structure of the underlying construct (Dimitrov, 2014). It aims to investigate the number and interconnections of latent factors that underlie the set of observable variables (i.e., items). Meanwhile, confirmatory factor analysis (CFA) is implemented to test the validity of a hypothesized model of latent factors and the factors' relationships to a set of observed variables (Dimitrov, 2014). The present paper will focus on confirmatory factor analysis of 6 alternative multivariate models that reflect different hypothetical factorial structures of the PTSD construct.

3.1. Models

Six models were considered for CFA (Figure 1). The assumptions underlying Model 1, Model 5, and Model 6 have been tested both during the development of the original version of ITQ by Cloître et al. (2018) and for the subsequent translations of the same measurement tool (e.g., Rossi et al, 2022). Unidimensional Model 1 is represented by a single latent factor PTSD on which all six core items of the PTSD subscale are loaded. This model is underlined by an alternative assumption that PTSD is a holistic, non-divisible construct that is manifested through a range of non-clustered symptoms. Model 2 is the second considered unidimensional model, but unlike Model 1, it governs 9 items - 6 core items and 3 functional impairment items. Models 3 and 4 are conceptualized as comprising two distinctive first-order latent factors: Trauma (T), on which 6 core items are loaded, and Severity (S), which governs 3 functional impairment items. In Model 3, T and S latent factors are correlated, whilst in Model 4 they are loaded on the second-order latent factor - PTSD. These two models aim to represent an alternative way to define and diagnose PTSD: instead of being characterized by 3 distinct symptom clusters, PTSD diagnostic requirements might include the general presence of trauma and functional impairment caused by it. Models 5 and 6 introduce three symptom clusters (Re-experiencing, Avoidance, and Sense of Threat) as latent factors, consistent with the initial theoretical formulation of PTSD within the ICD-11 diagnostic framework. In Model 5, three first-order latent factors representing symptom clusters (Re-experiencing, Avoidance, and

Sense of Threat) are governed by the second-order latent factor (PTSD). In Model 6, these latent factors are hypothesized to be only correlated. Thus, Model 5 outlines the hierarchical relationship between PTSD and its clinical features (i.e., Re-experiencing, Avoidance, and Sense of threat), while Model 6 proposes that as a construct, PTSD is not distinct from its clinical features but represented by their interrelationship. The underlying commonality of Model 1, Model 5, and Model 6 is the consideration of only core items that represent clinical and diagnostic features of PTSD with the exclusion of the items aimed at measuring the degree of functional impairment. To test for the relevance of the inclusion of functional impairment items into the multivariate model, Model 2, Model 3, and Model 4 were introduced. The goodness-of-fit of all six alternative models was tested separately for two samples (COVID wave 1 and COVID wave 2) and taken into account for the selection of the best-fitting model to reflect the factorial structure of ITQ's PTSD subscale.

Figure 1. Six alternative models of PTSD subscale's factorial structure.



3.2. Fit Statistics

The CFA was conducted using the maximum likelihood estimation method (ML) that assumes that the observed variables are continuous and normally distributed - as they resulted according to the items' descriptive statistics (Iacobucci, 2009). The fit of each alternative model was evaluated using goodness-of-fit (GOF) indices. GOF indices provide recommended cutoff values for assessing fit in SEM and are necessary to establish the validity of interpretations of the given model (Marsh, 2004). The following GOF indices were employed in the models' assessment: chi-square (χ^2, $p > .05$); Comparative Fit Index (CFI, ≥ .90); Tucker Lewis Index (TLI, ≥ .90); Root Mean Square Error of Approximation (RMSEA, < .05); Standardized Root Mean Square Residual (SRMR, <.08). Bayesian information criterion (BIC) and Akaike's information criterion (AIC) were implemented for the model comparison with lower relative values of BIC or AIC indicating better model fit. The values of the aforementioned goodness-of-fit indices of the 6 alternative models for two samples can be found in Table 1 and Table 2 respectively. Overall, χ^2 of all alternative models resulted in being statistically significant ($p < .001$). However, it should not lead to the rejection of the models: χ^2 is sensitive to sample size, and as sample size increases, χ^2 increases and its p-value decreases (Iacobucci, 2009). Thus, χ^2 tends to be significant (indicating a poor fit) with large sample sizes as in the case of the present study. Based on the results of CFA, Model 2 was discarded as it didn't reach acceptable values of CFI, TLI, and RMSEA. Despite meeting the cut-off threshold values for TLI, CFI, and SRMR, Model 4 was rejected due to computational infeasibility in estimating standard errors and inverting the information matrix, signaling model non-identifiability. Model 1 and Model 3 failed to reach the cut-off value of RMSEA (< .05) to prove the goodness-of-fit. Both Model 5 and Model 6 exhibited an excellent fit across both samples, as evidenced by the absolute values of CFI, TLI, RMSEA, and SRMR, coupled with the lowest relative BIC and AIC values. No differences in the values of considered GOF statistics were found between Model 5 and Model 6 as the same number of parameters were estimated in each model. In both models, all items loaded significantly positively onto latent factors representative of their respective symptom clusters, and all three latent factors (Re-experiencing, Avoidance, and Sense of Threat) loaded significantly positively on the second-order latent factor PTSD in Model 5 (Table 3 and Table 4).

Table 1. Goodness-of-fit indices of 6 alternative models of the PTSD subscale (COVID wave 1 sample, n=1048).

Model	*df*	χ^2	*p*	*CFI*	*TLI*	*AIC*	*BIC*	*RMSEA (90% CI)*	*SRMR*
1	9	237	< .001	.951	.918	15125.0	15184.3	.156 (.139 - .174)	.034
2	27	1063	< .001	.877	.837	21866.0	21955.0	.192 (.182 - .202)	.054
3	26	364	< .001	.960	.945	21169.4	21263.4	.112 (.102 - .122)	.034
4	25	364	< .001	.960	.942	21171.4	21270.4	.114 (.104 - .125)	.034
5	6	23	< .001	.996	.991	14916.5	14990.7	.052 (0.30 - .075)	.011
6	6	23	< .001	.996	.991	14916.5	14990.7	.052 (.030 - 075)	.011

Table 2. Goodness-of-fit indices of 6 alternative models of the PTSD subscale (COVID wave 2 sample, n=544).

Model	*df*	χ^2	*p*	*CFI*	*TLI*	*AIC*	*BIC*	*RMSEA (90% CI)*	*SRMR*
1	9	114	< .001	.960	.934	7453.8	7505.4	.147 (.123 - .171)	.033
2	27	578	< .001	.891	.854	10456.0	10533.4	.194 (.180 - .208)	.047
3	26	246	< .001	.956	.940	10125.8	10207.5	.125 (.111 - .139)	.035
4	25	246	< .001	.956	.937	10127.8	10213.8	.128 (.113 - .142)	.035
5	6	9	< .001	.999	.997	7354.5	7419.0	.031 (.000 - .069)	.006
6	6	9	< .001	.999	.997	7354.5	7419.0	.031 (.000 - .069)	.006

Table 3. Standardized factor loadings of Model 5 (COVID wave 1 sample, n = 1048).

Item	Re	Av	Th	PTSD
Upsetting dreams (Re1)	.838			
Reliving the event in the here and now (Re2)	.854			
Internal avoidance (Av1)		.907		
External avoidance (Av2)		.916		
Being on guard (Th1)			.821	
Feeling jumpy or easily startled (Th2)			.846	
Second-order factor loadings				
Re-experiencing				.964
Avoidance				.917
Sense of Threat				.938

Note: Re = Re-experiencing; Av = Avoidance; Th = Sense of Threat; PTSD = Posttraumatic Stress Disorder

Table 4. Standardized factor loadings of Model 6 (COVID wave 2 sample, n = 544).

Item	Re	Av	Th
Upsetting dreams (Re1)	.831		
Reliving the event in the here and now (Re2)	.884		
Internal avoidance (Av1)		.924	
External avoidance (Av2)		.940	
Being on guard (Th1)			.793
Feeling jumpy or easily startled (Th2)			.874

Note: Re = Re-experiencing; Av = Avoidance; Th = Sense of Threat; PTSD = Posttraumatic Stress Disorder

3.3. Model Comparison

The current CFA findings yield important insights. Firstly, the loading of functional impairment items onto latent factors representing either symptom cluster or PTSD resulted in insufficient GOF of the overall models (i.e., Model 2, Model 3, and Model 4). Thus, the hypothesized inclusion of these items in the factorial structure of the PTSD subscale shall be discarded. Secondly, the excellent fit of Model 5 and Model 6 creates continuity with the array of previous studies (Cloître et al., 2019; Ho et al., 2021; Rossi et al., 2022). Models 5 and 6 are the factorial substructures of the two-factor second-order model and correlated six-factor first-order model of the ITQ respectively. These two models, which reflect both PTSD and CPTSD subscales, were proven to fit the construct structure the best in the stage of the ITQ development and the further psychometric studies of ITQ translations. However, the choice of the superior model among these two has not been univocal. While the original research of Cloître et al. (2018) did not find any significant difference between second-order and first-order models in terms of their global and local fit, validation studies of ITQ translations gave preference to either first-order model (i.e., Model 6), as it is the case for the Chinese (Ho et al., 2021) and Italian (Rossi et al., 2022) translation validations, or to the second-order model (i.e., Model 5) as observed in French (Cyr et al., 2022) and Dari (Andisha et al., 2023) translation validation studies. In the present paper, no differences in fit emerged between Model 5 and Model 6. Nevertheless, preference shall be directed towards Model 5: a single-factor second-order model with three first-order latent factors. This selection is underpinned by the following theoretical considerations:

1) Despite shared core diagnostic criteria, PTSD and CPTSD are categorized as distinct stress-related disorders. Structuring symptom clusters under two separate subscales (i.e., second-order latent factors) aligns with the differential diagnosis practice for both disorders and their distinct classification within the ICD-11 framework (WHO, 2019).

2) Organizing ITQ in two separate subscales motivates and justifies the individual use of subscales in multiple research as in the case of the present study and validation study of the Danish translation of the ITQ (Hansen et al., 2021). Introducing second-order latent factors into the multivariate model enables the use of the PTSD subscale independently from the CPTSD subscale with the purpose of elucidating the presence of trauma in participants without necessitating a specific differentiation between PTSD and CPTSD. This approach would be less viable if the first-order six-factor model were implemented, as it would challenge the rationale behind the selection of symptom clusters for the assessment of the general presence of trauma.

Thus, in the subsequent analyses, Model 5 shall be regarded as the underlying factorial structure of the PTSD subscale.

4. Reliability

Reliability is a key concept in measurement and plays a pivotal role in evaluating the validity of assessment data (Dimitrov, 2014). Cronbach's coefficient alpha (α) has been implemented in nearly all prior ITQ validation research to measure the scale's internal consistency. ITQ generally demonstrated high internal consistency across the studies (α > .80), which aligns with the present paper's findings (α = .93). In addition to the aforementioned Cronbach's alpha, Mcdonald's omega (Ω), greatest lower bound (GLB), and composite reliability were computed. As well as Cronbach's alpha, all reliability statistics confirmed high internal consistency of PTSD subscale scores:

McDonald's omega (Ω) = .95

Greatest lower bound (GLB) = .94

Composite reliability =.97

To scrutinize more thoroughly the internal consistency of the measurement tool, item-total correlations for subscale items were computed. All item-total correlations were above the value of 0.8 across both samples, indicating good internal consistency of scores obtained from COVID wave 1 and wave 2 samples. The further step of reliability analysis concerned item-tem correlations. Pearson correlation coefficients were employed for the computation of item-item correlations, given that item distribution demonstrated a sufficient approximation to normal distribution. The latter observation was substantiated by skewness and kurtosis values falling within acceptable ranges for normality, specifically -1 to +1 and -2 to +2, respectively (Table 5). ITQ is a parsimonious and brief assessment tool that comprises only two items representing each symptom cluster of either PTSD or CPTSD. Item-item correlation test demonstrated the statistically significant correlation (r > .50) and the absence of redundancy among items. For the COVID wave 1 sample, items loading on the same latent factor (i.e., representing the same symptom cluster) had the strongest pairwise correlation compared to the rest of the items (Table 6). However, for the COVID wave 2 sample, the same observation didn't hold for all the items: for instance, Re2 displayed a slightly stronger correlation with Av1 and Av2, while Th2 showed a slightly stronger correlation with Av2 than with Th1. (Table 7).

Table 5. Item distribution for COVID wave 1 and wave 2 samples

	COVID wave 1 sample			COVID wave 2 sample		
	skew	kurtosis	se	skew	kurtosis	se
Re1	.92	- .31	.04	1.10	.14	.05
Re2	.70	- .68	.04	.87	- .35	.05
Av1	.65	- .62	.04	.83	- .36	.05
Av2	.71	- .51	.04	.81	- .50	.05
Th1	.46	- .85	.04	.61	- .65	.05
Th2	.74	- .51	.04	.88	- .32	.05

Table 6. Item-Item correlations (COVID wave 1 sample, n=1048)

	Re1	Re2	Av1	Av2	Th1	Th2
Re1	1.00					
Re2	.72	1.00				
Av1	.66	.70	1.00			
Av2	.67	.69	.83	1.00		
Th1	.61	.62	.65	.67	1.00	
Th2	.67	.65	.64	.66	.69	1.00

Table 7. Item-Item correlations (COVID wave 2 sample, n=544)

	Re1	Re2	Av1	Av2	Th1	Th2
Re1	1.00					
Re2	.73	1.00				
Av1	.69	.75	1.00			
Av2	.72	.75	.87	1.00		
Th1	.58	.61	.63	.66	1.00	
Th2	.63	.68	.67	.72	.69	1.00

5. Measurement Invariance

5.1 Conceptual Outline: Definition and Levels of Measurement Invariance

Measurement invariance (MI) is the property that ensures a test, scale, or other assessment tool functions the same way across different groups (cross-group measurement invariance) or time points (longitudinal measurement invariance) and that the construct is given the same meaning by those groups or across selected measurements (Putnick & Bornstein, 2016). Thus, a measure is invariant when members of different groups or populations who have the same standing on the construct receive the same observed score (Schmitt & Kuljanin, 2008). MI is one of the main prerequisites for meaningful score comparison across groups. Testing for MI has been necessitated by the increasing interest in comparative, intercultural, and international research. Within the structural equation framework (SEM), four different levels of MI can be defined (Putnick & Bornstein, 2016): configural, metric (weak factorial), scalar (strong factorial), and residual (strict factorial) invariance.

1) *Configural invariance* is the weakest form of factorial invariance and represents the first step in the hierarchy of MI testing. It is designed to test whether the construct has the same structure (i.e., number of latent factors and loading patterns) across tested groups (Hirschfeld & von Brachel, 2019).
2) *Metric invariance* is defined as the equivalence of the item loadings on the factor. By constraining factor loadings across the groups, it is possible to demonstrate whether each item of the scale contributes to the respective latent factor to the same extent across tested groups. Establishing metric invariance of the scale is required to allow for the comparison of the relationships between latent variables across the groups, such as genders, age groups, and cultural groups (Putnick & Bornstein, 2016).
3) *Scalar invariance* is tested by imposing equality constraints on item loadings and item intercepts across the groups. This step of MI implies that the meaning of the construct (i.e., factor loadings) and the levels of the underlying items (i.e., intercepts) are similar or equal across the groups. Once established, it enables the comparison of the scores on the latent variables between the groups (Van De Schoot et al., 2012).
4) *Residual invariance* is the ultimate level in the MI hierarchy and indicates the invariant uniqueness of the measurement. It is tested by constraining factor loadings, intercepts, and residuals to be equal across the groups, and it is aimed to illustrate whether the latent construct is measured identically across the groups (Van De Schoot et al., 2012).

Thus, the process of establishing MI is systematic as the levels of MI are organized into a hierarchical structure. Each subsequent level in this hierarchy imposes progressively stricter equality constraints on the factorial structure. This step-wise procedure ensures a comprehensive investigation of the measure's functioning across different groups. If measurement non-invariance is observed at any particular level, establishing higher levels of MI becomes not feasible. However, when non-invariance is detected, it is necessary to either revise the construct of interest to ensure that it is not inherently non-invariant or to investigate the source of non-invariance by releasing or adding the constraints to retest the redefined model on the lower levels of MI and to establish partial invariance.

5.2 Testing for Measurement Invariance: Multiple-Group CFA

Multiple-Group Confirmatory Factor Analysis (MG-CFA) is one of the most widespread statistical techniques used to investigate the degree to which the measures are invariant across groups. Establishing MI via MG-CFA involves testing a series of hierarchical and progressively constrained measurement models. Equality constraints imposed on the sequence of nested models reflect the requirements for each level of the measurement invariance hierarchy (Hirschfeld & von Brachel, 2019). In total, four models are evaluated and compared in a pairwise manner:

1) Model 1a tests for configural invariance - no equality constraints are imposed on the factorial structure of the model, and the MG-CFA consists of estimating the GOF and the significance of the loadings of the model across the groups.
2) Model 2a tests for metric invariance - factor loadings are constrained to be equal across the groups. The model's fit is estimated via the assessment of GOF indices values and compared to the fit of the baseline model, i.e., Model 1a.
3) Model 3a tests for scalar invariance - factor loadings and intercepts are constrained to be equal across the groups. The absolute and relative fit is evaluated in order to establish scalar invariance.
4) Model 4a tests for residual invariance - factor loadings, intercepts, and residuals are constrained to be equal across tested groups. The constrained model is fitted to the data of both groups and estimated based on the absolute values of the GOF indices and compared to Model 3a.

MG-CFA assesses the change in fit indices when cross-group constraints are imposed on each invariance level. The following GOF indices were considered for the estimation of the fit of the constrained models: non-significant χ^2 ($p > .05$), CFI (> .95), TLI (> .95), RMSEA (< .05), SRMR (< .05) (Van De Schoot et al., 2012). Configural invariance is established if the values of GOF

indices point out an excellent fit of the baseline model across the groups and if the same loadings are significant. However, to establish metric, scalar, and residual invariance, comparative fit estimation of pairs of nested models is required. Metric invariance holds if the fit of Model 2a is not significantly worse than the fit of Model 1a; scalar invariance is established if the fit of Model 3a is not substantially worse than the overall fit of Model 2a; residual invariance is supported if the fit of 4a Model is not significantly worse than the one of Model 3a (Putnick & Bornstein, 2016). The use of different rules to judge whether the decrease in model fit is substantial or acceptable is an open debate. Classically, MI was evaluated using only the significance of $\Delta\chi^2$ in the comparison of nested models. However, due to the sensitivity of the χ^2 to insignificant deviations within large samples, the focus was shifted to alternative fit indices that are less sensitive to sample size (e.g., ΔCFI, ΔRMSEA, and ΔSRMR) (Putnick & Bornstein, 2016). The present paper took up a mixed approach, and to allow for the comparison of the fit between nested models, the difference in 3 indices was evaluated: non-significance of $\Delta\chi^2$ ($p > .05$), ΔCFI (< .010), and ΔRMSEA (< .015) (Hirschfeld & von Brachel, 2019). Metric, scalar, or residual invariance was supported if at least two out of three index differences complied with cut-off thresholds (Cheung & Rensvold, 2002).

5.3 Prior Evidence for the ITQ's Measurement Invariance

The validity of the ITQ has been investigated extensively in an array of prior research, covering both the original version of the questionnaire and its subsequent translations to other languages. However, few studies tested the cross-group MI of the scale and no studies were conducted to test longitudinal measurement invariance. In the original paper, Cloître et al. (2021) reported the evidence for the ITQ's configural and metric MI across clinical and community samples. The model with configural invariance had an acceptable model fit. Metric invariance of the scale was supported by insignificant differences in CFI and RMSEA. However, $\Delta\chi^2$ appeared to be statistically significant ($\Delta\chi2 = 13.97$, $\Delta df = 6$, $p = .030$), suggesting a substantially worse fit of the constrained model (Cloître et al, 2018). The evidence for the absence of differential item functioning (DIF) is rather inconsistent. Cloître et al. (2018) found no DIF for any ITQ item across community and clinical samples. On the other hand, Nielsen et al. (2023), in a study conducted on a multicultural clinical refugee sample, demonstrated the presence of DIF for two PTSD subscale items relative to gender and time since the occurrence of the traumatic event and found no DIF relative to language groups (Danish, Arabic, and Bosnian) and the level of interpreter-assisted administration.

5.4 Measurement Invariance of the ITQ's PTSD Subscale

The present paper focused on the assessment of the MI of the PTSD subscale across two time points: COVID wave 1 and COVID wave 2. Four levels of MI were tested via MG-CFA in the following logical order resulting from the definition of MI: configural, metric, scalar, and residual. Based on the evidence provided by the aforementioned CFA of several alternative multivariate models, the single-factor second-order model with three first-order factors (i.e., Model 5) was chosen as a baseline model of the PTSD subscale for MI testing.

Configural invariance. Model 1a demonstrated high, above-cut-off values of GOF indices and consequently an excellent model fit (Table 8). Due to the large sample size, χ^2 appeared to be statistically significant, but as discussed previously, it should not lead to the rejection of the baseline model and the configural invariance of the PTSD subscale.

Metric invariance. Imposing equality constraints on factor loadings across COVID wave 1 and COVID wave 2 samples did not produce a substantial worsening of the model fit. Instead, Model 2a had an excellent model fit with no observed worsening in CFI, an insignificant improvement in TLI and RMSEA indices, and an increase in SRMR (Table 8). Pairwise comparison of Model 1a and Model 2a extended the evidence for the goodness-of-fit of the model: $\Delta\chi^2$ was statistically insignificant ($p > .1245$), while ΔCFI and ΔRMSEA were below cutoff values (Table 9). Thus, the absolute and relative model fit of Model 2a indicated metric invariance of the PTSD subscale.

Scalar invariance. Introducing equality constraints on factor loadings and intercepts across the samples did not significantly influence the fit of the resulting Model 3a. The values of GOF indices remained within cut-off thresholds with a minor increase in TLI and SRMR values and a decrease in RMSEA (Table 8). The presence of scalar invariance is further evidenced by the statistical insignificance of $\Delta\chi^2$, ΔCFI, and ΔRMSEA values (Table 9).

Residual invariance. Model 4a, bearing equality constraints imposed on factor loadings, intercepts, and residuals, demonstrated good model fit (Table 9). The values of GOF indices of the model did not exceed the respective cut-offs, however, within-limit worsening was found for all the considered indices, with the largest increase in RMSEA value (ΔRMSEA = .007). The chi-square difference test (DIFFTEST) of Model 3a and Model 4a showed that $\Delta\chi^2$ was statistically significant ($p = .00027$) which indicates potential residual non-invariance (Table 9). To investigate the source of non-invariance, item residuals of Model 4a were inspected in order to understand which residual variances were different across the groups. The conducted analysis (i.e., Lagrange Multiplier test) suggested that releasing the constraints from the residuals of the Re2 ("Reliving the event in the here in now"), Av1 ("Internal avoidance"), and Av2 ("External avoidance") items would improve

the overall model fit and minimize the difference between Model 3a and Model 4a. To test this hypothesis, Model 4b was devised and juxtaposed with Model 3a. In Model 4b, residual variances of Re2, Av1, and Av2 were allowed to vary, while factor loadings, intercepts, and the remaining item residuals were constrained to be equal across the groups. The overall fit of Model 4b was proved to be better than the fit of Model 4a relative to the values of GOF indices, and the pairwise comparison of Model 3a and Model 4b demonstrated that releasing target constraints resulted in insignificant $\Delta\chi^2$ and lower ΔCFI and ΔRMSEA. However, releasing the constraints from the residuals of the items Re2, Av1, and Av2 and establishing partial residual invariance does not seem strictly necessary. Firstly, within the context of the PTSD subscale and its brevity, releasing residual constraints from three items would mean imposing residual equality constraints only on half of the items, which is a substantial model change and an overstretch of the invariance assumption. Secondly, the significance of the $\Delta\chi^2$ is not necessarily symptomatic of the measurement non-invariance. As outlined previously in the paper, alternative fit indices (AFI), such as ΔCFI and ΔRMSEA, play a role in indicating MI of the scale. As ΔCFI and ΔRMSEA between Model 3a and Model 4a remained within acceptable cut-off thresholds (< .010 and < .015 respectively), it is possible to assume that residual invariance of the PTSD subscale does hold across COVID wave 1 and COVID wave 2 samples despite the significance of the $\Delta\chi^2$. Thus, upon considerations mentioned above, the preference shall be given to the model with fully constrained residual variances, that is to Model 4a, and to the assumption of the full residual invariance of the ITQ's PTSD subscale.

Table 8. Results of Multiple-Group CFA of 4 +1 nested models

Model	χ^2	*df*	*p*	*CFI*	*TLI*	*RMSEA (90% CI)*	*SRMR*	*AIC*	*BIC*
1a	31.85	12	0.001	.997	.993	.046 (.027 - .065)	.008	22295	22520
2a	40.49	17	0.001	.997	.994	.042 (.025 - .059)	.020	22294	22492
3a	43.22	19	0.001	.997	.995	.040 (.024 - .056)	.021	22292	22480
4a	68.78	25	0.000	.994	.993	.047 (.034 - .060)	.021	22305	22461
4b	44.74	22	0.003	.997	.996	.036 (.021 - .051)	.020	22287	22459

Table 9. Pairwise comparison of 4 + 1 nested models

Models	*Δdf*	*Δχ²*	*p*	*ΔCFI*	*ΔRMSEA*
2a - 1a	5	8.64	.1245	.000	-.004
3a - 2a	2	2.73	.2555	.000	-.002
4a - 3a	6	25.56	**.00027**	- .003	.007
4b - 3a	3	1.53	**.6752**	.000	.-004

6. Discussion

The present paper contributes to the existing body of psychometric research on the ITQ by examining the construct validity of its PTSD subscale. The aforementioned results of the CFA supported the evidence in favor of a single-factor second-order model with three first-order latent factors as the underlying factorial structure of the PTSD subscale (i.e., Model 5). These findings create continuity with previously conducted research in which the extended version of the same measurement model was tested and selected as the best-fitting one (e.g., Cloître, 2018). Excellent fit of the hierarchical model of the PTSD subscale reflects the definition and clinical picture of PTSD as outlined in the ICD-11 (WHO, 2019): PTSD is a complex construct that governs three symptom clusters that characterize the disorder and indicate the presence of experienced trauma. As the Italian version of the ITQ was used in this study, the results support the overall structural validity of the ITQ's PTSD subscale as well as the Italian translation of the questionnaire. Reliability analysis of the PTSD subscale showed that the measure has great internal consistency according to several statistics (α = .93, Ω = .95, GLB = .94). To unite reliability analysis with CFA findings, composite reliability of the scale was computed based on the factorial structure of Model 5. Once again, the high value of composite reliability (= .97) supported the high internal consistency of the PTSD scale.

The ITQ is a concise and straightforward self-report questionnaire that is valuable both in research and practice. It enables differential diagnosis between PTSD and CPTSD, and importantly, helps identify the presence of trauma in clinical and subclinical cases. Since its development and primary validation in 2018 by Cloître et al., the ITQ has been used in numerous psychotraumatology studies across various populations (e.g., community and clinical samples) and cultures. The necessity to assess and compare results from different groups raised the question of the tool's measurement invariance. The present study conducted MI testing which yielded significant results. By

implementing MG-CFA, it was demonstrated that ITQ's PTSD subscale is invariant on configural, metric, and scalar levels of measurement invariance. The evidence for the scale's scalar invariance is crucial for further research as it is a main prerequisite that enables latent mean comparison across the groups. The results on residual invariance of the measure were less definitive. Alternative fit indices (AFI), such as ΔCFI and ΔRMSEA, supported residual invariance, however, the $\Delta\chi^2$-test of nested Model 3a and Model 4a contradicted this assumption as it was statistically significant. Despite being an obligatory step for establishing full uniqueness MI, residual invariance is not a prerequisite for comparing the groups on their scores on the latent variables as the item residuals do not contitute a part of the latent factor (Putnick & Bornstein, 2016). Establishing residual invariance only suggests that the scale measures the latent construct across the groups or time points identically, i.e., with the same degree of measurement error (Van De Schoot et al., 2012). The present paper utilized a mixed approach, considering both AFIs and $\Delta\chi^2$. Based on the statistical insignificance of ΔCFI and ΔRMSEA, residual invariance of the PTSD subscale was suggested (Cheung & Rensvold, 2002).

The main limitation of this study is its generalizability. Given that the analysis focused solely on the PTSD subscale of the ITQ, it does not provide evidence for the longitudinal measurement invariance of the entire questionnaire. The longitudinal measurement invariance of the DSO subscale and the full ITQ are yet to be confirmed. Despite this, the examination of the PTSD subscale's structural validity enables its individual use in future studies where a differential diagnosis of PTSD and CPTSD is not required.

7. Conclusions

This study scrutinized the construct validity of the ITQ's PTSD subscale by assessing its factorial structure and longitudinal MI. The hierarchical, second-order factorial structure of the PTSD construct, developed and confirmed in the previously conducted research, was unequivocally supported by the present CFA analysis. Testing for longitudinal MI yielded results in favor of the scale's MI on configural, metric, scalar, and residual levels. The overall findings of the study suggest that ITQ's PTSD subscale is a valuable measurement tool that allows for the assessment of symptoms associated with PTSD in a reliable and context-insensitive manner. However, it's important to recognize the limitations of this study. Future research could focus on extending the evidence for cross-group and longitudinal MI by including both ITQ's subscales and incorporating more diverse groups.

8. Bibliography

1. Andisha, P., Shahab, M. J., & Lueger‐Schuster, B. (2023). Translation and validation of the Dari International Trauma Questionnaire (ITQ) in Afghan asylum seekers and refugees. *European Journal of Psychotraumatology*, *14*(1)
2. Bruno, G., Panzeri, A., Granziol, U., Alivernini, F., Chirico, A., Galli, F., Lucidi, F., Spoto, A., Vidotto, G., & Bertamini, M. (2020). The Italian COVID-19 Psychological Research Consortium (IT C19PRC): General Overview and Replication of the UK Study. Journal of Clinical Medicine, 10(1), 52.
3. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.
4. Cloître, M., Hyland, P., Prins, A., & Shevlin, M. (2021). The International Trauma Questionnaire (ITQ) measures reliable and clinically significant treatment-related change in PTSD and complex PTSD. *European Journal of Psychotraumatology, 12*(1).
5. Cloître, M., Shevlin, M., Brewin, C. R., Bisson, J. I., Roberts, N. P., Maercker, A., Karatzias, T., & Hyland, P. (2018). The International Trauma Questionnaire: development of a self‐report measure of ICD‐11 PTSD and complex PTSD. *Acta Psychiatrica Scandinavica, 138(6)*, 536–546.
6. Cyr, G., Bélanger, C., & Godbout, N. (2022). French translation and validation of the International Trauma Questionnaire in a Canadian community sample. *Child Abuse & Neglect, 128*, 105627.
7. Del Re, D., Palla, L., Meridiani, P., Soffi, L., Loiudice, M. T., Antinozzi, M., & Cattaruzza, M. S. (2024). The spread in time and space of COVID-19 pandemic waves: the Italian experience from mortality data analyses. *Frontiers in Public Health, 12*.
8. Dimitrov, D. M. (2014). *Statistical methods for validation of assessment scale data in counseling and related fields*.
9. Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research. *Social Psychological and Personality Science, 8*(4), 370–378.
10. Hansen, M., Vægter, H. B., Cloître, M., & Andersen, T. E. (2021). Validation of the Danish International Trauma Questionnaire for posttraumatic stress disorder in chronic pain patients using clinician-rated diagnostic interviews. *European Journal of Psychotraumatology, 12(1)*.
11. Ho, G. W., Karatzias, T., Cloître, M., Chan, A. C. Y., Bressington, D., Chien, W. T., Hyland, P., & Shevlin, M. (2019). Translation and validation of the Chinese ICD-11 International Trauma Questionnaire (ITQ) for the Assessment of Posttraumatic Stress Disorder (PTSD) and Complex PTSD (CPTSD). *European Journal of Psychotraumatology, 10(1)*.

12. Hirschfeld, G., & von Brachel, R. (2019). Improving Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research, and Evaluation,* 19(7)
13. Hyland, Shevlin, Brewin, Cloitre, Downes, Jumbe, Karatzias, Bisson, & Roberts. (2017). Validation of post-traumatic stress disorder (PTSD) and complex PTSD using the International Trauma Questionnaire. *Acta Psychiatrica Scandinavica, 136(3), 313–322*
14. Iacobucci, D. (2009). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90–98.
15. Marsh, H. W., Hau, K. -T., & Wen, Z. (2004). In search of golden rules. *Structural Equation Modeling, 11*(3), 320−341.
16. Nielsen, T., Elklit, A., Vang, M. L., Nielsen, S. B., Auning-Hansen, M., & Palić, S. (2023). Cross-cultural validity and psychometric properties of the International Trauma Questionnaire in a clinical refugee sample. *European Journal of Psychotraumatology*, *14*(1)
17. 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.
18. R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/.
19. Redican, E., Nolan, E., Hyland, P., Cloître, M., McBride, O., Karatzias, T., Murphy, J., & Shevlin, M. (2021). A systematic literature review of factor analytic and mixture models of ICD-11 PTSD and CPTSD using the International Trauma Questionnaire. *Journal of Anxiety Disorders, 79*, 102381.
20. Rossi, R., Socci, V., Pacitti, F., Carmassi, C., Rossi, A., Di Lorenzo, G., & Hyland, P. (2022). The Italian version of the International Trauma Questionnaire: Symptom and Network Structure of Post-Traumatic Stress Disorder and Complex Post-Traumatic Stress Disorder in a sample of late adolescents exposed to a natural disaster. *Frontiers in Psychiatry, 13*.
21. Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222.
22. Van De Schoot, R., Lugtig, P., & Hox, J. J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740
23. Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, *34*(5), 719–751.

24. World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/