A validation study of the Dutch Childhood Trauma Questionnaire-Short Form: factor structure, reliability, and known-groups validity

Brett D. Thombs¹; David P. Bernstein²; Jill Lobbestael²; Arnoud Arntz²

¹Department of Psychiatry, McGill University and Sir Mortimer B. Davis – Jewish General Hospital, Montreal, Quebec, Canada; ²Department of Clinical, Medical, and Experimental Psychology, University of Maastricht, Maastricht, the Netherlands;

Address for Correspondence:
Brett D. Thombs, Ph.D.
Institute of Community and Family Psychiatry
SMBD-Jewish General Hospital
4333 Cote Ste Catherine Road
Montreal, Quebec H3T 1E4
Tel (514) 340-8222 ext. 5112
Fax (514) 340-8124
E-mail: brett.thombs@mcgill.ca

Dr. Thombs is supported by a New Investigator Award from the Canadian Institutes of Health Research and an Établissement de Jeunes Chercheurs award from the Fonds de la Recherche en Santé Québec.
Abstract

Objective: The 28-item Childhood Trauma Questionnaire – Short Form (CTQ-SF) has been translated into at least 10 different languages. The validity of translated versions of the CTQ-SF, however, has generally not been examined. The objective of this study was to investigate the factor structure, internal consistency reliability, and known-groups validity of the Dutch CTQ-SF.

Methods: A total of 261 patients receiving mental health treatment and 227 non-clinical controls were administered the Dutch CTQ-SF. Confirmatory factor analysis (CFA) was done with Mplus for ordinal data to test the 5-factor structure of the CTQ-SF. Cronbach’s alpha was computed, and known-groups validity was assessed by comparing CFA latent factor levels between clinical and non-clinical respondents.

Results: The 5-factor model fit well, although one item (I believe I was molested) was removed due to high levels of missing data and because it loaded on the Physical Abuse factor rather than the intended Sexual Abuse factor. Cronbach’s alpha was 0.91 for Physical Abuse, 0.89 for Emotional Abuse, 0.95 for Sexual Abuse, 0.63 for Physical Neglect, and 0.91 for Emotional Neglect. Latent factor levels were more than one standard deviation higher (p < .001) for patients receiving mental health treatment than controls for all CTQ-SF five scales.

Conclusions: The findings from this study provide evidence for the validity and reliability of the 24-item Dutch CTQ-SF, but showed that one item translated from the 25-item English CTQ-SF, I believe I was molested, was not a valid indicator of childhood sexual abuse in the Dutch version and should not be used.

Practice Implications: Researchers and clinicians should use the 24-item version of the Dutch CTQ-SF.
Introduction

The Childhood Trauma Questionnaire-Short Form (CTQ-SF) is a 28-item retrospective self-report questionnaire designed to assess five dimensions of childhood maltreatment: (1) physical abuse, (2) emotional abuse, (3) sexual abuse, (4) physical neglect, and (5) emotional neglect. A March 2009 MEDLINE search found 141 references to the CTQ-SF, whereas no other self-report measure of childhood maltreatment listed in a recent review of childhood trauma assessment tools (Roy & Perry, 2004) was referenced more than 15 times. The CTQ-SF (Bernstein & Fink, 1998; Bernstein et al., 2003) has been translated into German (Bader, Schafer, Schenkel, Nissen, & Schwander, 2007; Driessen et al., 2000; Driessen, Schroeder, Widmann, von Schonfeld, & Schneider, 2006; Gast, Rodewald, Nickel, & Emrich, 2001; Kersting et al., 2003; Krischer & Sevecke, 2008; Moehler, Biringen, & Poustka, 2007; Schafer et al., 2006; Schafer et al., 2007; Spitzer, Barnow, Gau, Freyberger, & Grabe, 2008; Woller, Hartkamp, & Tress, 2007), Norwegian (Fosse & Holen, 2002; Fosse & Holen, 2006; Fosse & Holen, 2007), Turkish (Aslan & Alparslan, 1999; Mirsal, Kalyoncu, Pektas, Tan, & Beyazyurek, 2004; Sar, Akyuz, Kundakci, Kiziltan, & Dogan, 2004; Sar, Akyuz, Kugu, Ozturk, & Ertem-Vehid, 2006; Sar, Unal, & Ozturk, 2007; Ucok & Bikmaz, 2007), Spanish (Basurte, Diaz-Marsa, Martin, & Carrasco, 2004), French (Collin-Vezina, Cyr, Pauze, & McDuff, 2005; Perroud et al., 2008), Haitian Creole (Martolf, 2004), Portuguese (Grassi-Oliveira, Stein, & Pezzi, 2006), Italian (Sarchiapone, Carli, Cuomo, & Roy, 2007), and Dutch (Arntz, 1999; Arntz, Dietzel, & Dreessen, 1999; Arntz, Meeren, & Wessel, 2002; Giesbrecht, Merckelbach, Kater, & Sluis, 2007; Lobbestael, Arntz, & Bernstein, under review; Lobbestael, Arntz, Harkema-Schouten, & Bernstein, under review; Wessel, Meeren, Peeters, Arntz, & Merckelbach, 2001), but translated versions have generally not been validated.
The Dutch CTQ-SF has been found to have good internal consistency in a small sample of 16 patients with borderline personality disorder, 12 patients with a cluster-C personality disorder, and 15 control subjects (Cronbach’s alpha ≥ 0.87 for all 5 scales), and patients with borderline personality disorder scored higher than other subjects on all scales (Arntz et al., 1999; Arntz, 1999). In addition, scores on the 5 CTQ-SF scales among Dutch respondents showed good convergent validity with responses from a semi-structured interview for childhood trauma (Lobbestael, Arntz, Harkema-Schouten et al., under review). No studies, however, have tested the structural validity of the translated Dutch CTQ-SF. The objective of this study was to test the factor structure, internal consistency reliability, and known-groups validity of a Dutch translation of the CTQ-SF in a large sample of clinical and non-clinical respondents.

Methods

Sample Selection and Procedure

The CTQ-SF was administered to participants in studies related to personality disorders, the association of child maltreatment history with other psychiatric disorders, and the validation of a childhood trauma interview at psychiatric hospitals, outpatient mental health clinics, forensic clinics and prisons in the Netherlands from 2004 to 2007. For all studies, individuals had to be between 18 to 70 years of age. Exclusion criteria included alcohol or drug intoxication, an IQ of < 80, and the inability to independently complete the CTQ-SF. In addition, non-patient controls were recruited via advertisements, flyers, or personal contacts through the researchers. Written informed consent was obtained from all participants after a full explanation of study procedures. Participants were compensated 10 euro for completing the procedure. The protocols for each study were approved by the Medical Ethical Committee of the Academic Hospital in Maastricht.
There are 5 items on each scale of the CTQ-SF (Bernstein & Fink, 1998; Bernstein et al., 2003) plus an additional 3-item minimization/denial scale. Item response categories are scored from 1-5 and are structured to reflect the frequency of maltreatment experiences (never true, rarely true, sometimes true, often true, very often true). Bernstein et al. (2003) used confirmatory analysis to demonstrate the validity of the 5-factor structure of the original CTQ-SF and reported good internal consistency reliability for each of the CTQ-SF scales across 4 heterogeneous clinical samples from the USA: Physical Abuse = .83 to .86, Emotional Abuse = .84 to .89, Sexual Abuse = .92 to .95; Physical Neglect = .61 to .78, Emotional Neglect = .85 to .91. Scher et al. (2001) also reported adequate to good internal consistency reliability in a large community sample from the USA: Physical Abuse = .69, Emotional Abuse = .83, Sexual Abuse = .94, Physical Neglect = .58, Emotional Neglect = .85. Reports on traumatic events on the CTQ-SF scales have good stability over time and good convergent and divergent validity when trauma histories have been ascertained via other methods, such as clinical interviews or independent case records (Bernstein, Fink, Handelsman, & Foote, 1994; Bernstein, Ahluvalia, Pogge, & Handelsman, 1997; Fink, Bernstein, Handelsman, & Foote, 1995; Lipschitz, Bernstein, Winegar, & Southwick, 1999). The Dutch CTQ-SF was generated using standard translation and back-translation procedures to ensure linguistic equivalence (Brislin, 1970).

Data Analyses

Patient sociodemographic and clinical characteristics were compared between clinical and non-clinical respondents. Differences between the groups were assessed using t tests for continuous data and the $\chi^2$ statistic or Fisher’s Exact test for categorical data. Confirmatory factor analysis (CFA) models were conducted with Mplus (version 3.11) (Muthén & Muthén, 1998-2004) to test the validity of the 5-factor structure of the CTQ-SF and to test the fit of
individual items on each factor. Item responses were ordinal Likert data, so the weighted least squares estimator with a diagonal weight matrix and robust standard errors and a mean- and variance-adjusted chi-square statistic (WLSMV) was used with delta parameterization (Muthén & Muthén, 1998-2004). Mplus procedures for estimating models that included patients with missing data were used. Modification indices were used to identify pairs of items within scales for which model fit would improve if error estimates were freed to covary, and for which there appeared to be theoretically justifiable shared method effects (McDonald & Ringo Ho, 2002). A chi-square goodness-of-fit test and 3 fit indices were used to assess model fit, including the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), the comparative fit index (CFI; Bentler, 1990), and the root mean square error of approximation (RMSEA; Steiger, 1990). Since the chi-square test is highly sensitive to sample size and can lead to the rejection of well-fitting models, practical fit indices were emphasized (Reise, Widaman, & Pugh, 1993). Guidelines proposed by Hu and Bentler (1999) suggest that models with TLI and CFI close to .95 or higher, RMSEA close to .06 or lower and SRMR close to .08 or lower are representative of good fitting models. A CFI of .90 or above (Kline, 2005) and a RMSEA of .08 or less (Browne & Cudeck, 1993), however, are also considered to represent reasonably acceptable model fit.

Cronbach’s alpha was computed for each CTQ-SF scale to assess internal consistency reliability. Known-groups validity was assessed by comparing latent factor levels in the CFA model for each scale between patients receiving mental health services and non-clinical controls.

Results

Sample Characteristics

Demographic, clinical, and maltreatment characteristics of study participants are shown in Table 1. There were 36 patients from psychiatric hospitals (7.4%), 178 from outpatient mental
health clinics (36.5%), 48 from forensic clinics and prisons (9.9%), and 227 non-patients (46.5%). There were 215 patients with at least one Axis I diagnosis (44.1%), including 80 with one (16.4%), 77 with two (15.8%), 47 with three (9.6%), 11 with four or five, (2.2%). Clinical and non-clinical respondents differed significantly on sex, nationality, marital status, education, occupational status, and mental health variables.

**CTQ-SF Item Responses and Factor Structure**

Of the 488 respondents, 447 respondents provided complete CTQ-SF responses with no missing data. Of the 41 respondents with any missing data, 33 did not respond to item SA21 (*I believe I was molested*), which is part of the Sexual Abuse scale in the original English CTQ-SF.

The standard five-factor model of the English CTQ-SF was tested initially. Three pairs of item error covariances were freed based on modification indices. In each case, both members of the pair were from the same scale and demonstrated shared method or format features. Consistent with the English CTQ-SF (Thombs, Lewis, Bernstein, Medrano, & Hatch, 2007), error variances were freed to covary for items EA3 and EA13, PN1 and PN6, and EN12 and EN17 (see Table 2). Modification indices also showed that item SA21 loaded more strongly on the Physical Abuse factor (0.66) than the Sexual Abuse factor (0.22). Based on this result and the high number of respondents with missing data on item SA21 a five-factor was estimated without SA21, leaving a total of 24 items, including 4 on the Sexual Abuse scale and 5 on all other scales. This resulted in a well-fitting model $\chi^2(59) = 163.2$, CFI = .99, TLI = .99, RMSEA = .06] in which all factor loadings on all scales were 0.55 or greater. Factor intercorrelations ranged from 0.58 to 0.88. Factor loadings are shown in Table 2.

**Reliability**
Cronbach’s alpha for the CTQ-SF scales was 0.91 for Physical Abuse, 0.89 for Emotional Abuse, 0.95 for Sexual Abuse, 0.63 for Physical Neglect, and 0.91 for Emotional Neglect.

Known-Groups Validity

To test the known-groups validity of the Dutch CTQ-SF scales, the final CFA model for the 24-item Dutch CTQ-SF was rerun with each factor regressed on clinical status (patients receiving mental health treatment versus controls). The model fit well \( \chi^2(67) = 174.6, \ CFI = .99, \ TLI = .99, \ RMSEA = .06 \). Latent factor levels were more than one standard deviation higher (\( p < .001 \)) for patients receiving mental health treatment than controls for all five scales: Physical Abuse (1.09 standard deviations), Emotional Abuse (1.41), Sexual Abuse (1.18), Physical Neglect (1.03), Emotional Neglect (1.04).

Discussion

The main findings of this study were (1) that the hypothesized five-factor model of the CTQ-SF fit the 24-item Dutch version of the CTQ-SF well, although one item (SA21 – I believe I was molested) was removed from the Sexual Abuse scale; (2) that the Dutch CTQ-SF effectively differentiated between patients receiving mental health treatment and non-clinical controls; and (3) that the five scales of the Dutch CTQ-SF demonstrated adequate internal consistency reliability that was similar to previous studies of the English version of the CTQ-SF, although, as in the English version, Cronbach’s alpha was somewhat lower for the Physical Neglect scale (Bernstein et al., 2003; Scher et al., 2001).

The item I believe I was molested (SA21) was removed from the Dutch version of the CTQ-SF. Whereas the English word molested clearly refers to sexual abuse in the context of a childhood maltreatment questionnaire, the meaning of the translated Dutch word gemolesteerd does not have sexual connotations. A major Dutch-English translation dictionary translates the
Dutch *molesteren* into English as meaning “annoy,” “beat up,” “wreck,” or “ruin” (Martin, Tops, Broeders, Roos, & Schrama, 1991). Results from this study suggest that many patients were not clear about the meaning of this item or interpreted it as related to physically abusive behavior. Due to the ambiguity of item SA21, it should not be used in the Dutch CTQ-SF.

One implication of removing item SA21 from the Dutch CTQ-SF is that the Sexual Abuse scale has 4 items, whereas the English CTQ-SF Sexual Abuse scale has 5 items. One potential concern is that this could limit the comparability of scores between versions. There are published cutoff scores from the English CTQ-SF, for instance, that can be used to identify patients most likely to have a history of sexual abuse, and one might wonder how these would be applied to the 4-item Sexual Abuse scale of the Dutch CTQ-SF. Cutoff points, however, developed for screening in one population group, would not necessarily be expected to be accurate in another group even with a validated instrument translation. Comparison of cutoff scores across groups requires the demonstration *measurement invariance* across groups or independent validation of cutoff scores in each group (Hui & Traindis, 1985; Poortinga, 1989). Based on the findings from this study, the most valid approach for comparison purposes would be to compare the 4-item Sexual Abuse scale between English- and Dutch-speaking groups and to develop Dutch cutoff scores based on the 4-item Sexual Abuse scale.

There are limitations that should be considered in interpreting results from this study. First, the sample was comprised of patients and controls from a series of other studies, so it is possible that results may not generalize fully to other community or patient samples. Related to this, clinical and non-clinical respondents to the CTQ-SF differed demographically and clinically, which also limits external validity. Second, as is the case in most retrospective studies
of childhood maltreatment, data was limited to self-report, and the concordance with actual histories of maltreatment was not confirmed.

In summary, the findings from this study provide initial evidence for the validity and reliability of the 24-item Dutch CTQ-SF, but showed that one translated item, *I believe I was molested*, was not a valid indicator of childhood sexual abuse in the Dutch version and should not be used. The Dutch CTQ-SF had adequate internal consistency reliability that was consistent with results from prior studies of the English CTQ-SF and effectively discriminated between clinical and non-clinical samples. Additional validation studies are needed that include corroboration with clinical interviews or case charts.
References


Lobbestael, J., Arntz, A., & Bernstein, D. Disentangling the relationship between different types of childhood trauma and personality disorders. Under review.


