The Harvard Trauma Questionnaire:
Reliability and Validity Generalization Studies of the Symptom Scales

Chantal Darzi

Dissertation submitted to the
School of Psychology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Clinical Psychology

Faculty of Social Sciences
University of Ottawa
Ottawa, Ontario

© Chantal Darzi, Ottawa, Canada, 2017

For Raja

et nos deux princesses,

Sophia et Mina

Abstract

The cross-cultural applicability of the PTSD diagnosis has been widely disputed in recent years. Consequently, an examination of the psychometric properties of instruments that are used to assess traumatized individuals of various cultures is of utmost importance. To respond to this need, the overall goal of this dissertation was to evaluate the psychometric properties of the Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992), a measure that was developed to assess trauma symptoms across cultures. In the first study, I conducted a search of all publications and dissertations that used the symptoms scales of the HTQ. This search revealed that the HTQ is commonly used by trauma researchers, however only a minority of them reported using established translation and cultural adaptation procedures to adapt the instrument for their specific sample. In addition, of the 384 studies considered for inclusion, only 44% of them reported internal consistency estimates of their sample. I then performed reliability generalization analyses on Cronbach's alpha coefficients to assess the reliability properties of the HTQ symptom scales. Overall, 103 samples were included in the analyses, representing various cultures, languages and countries of study. The findings of this study indicated that both the HTQ-16 and 30 symptom scales are likely to provide reliable scores across diverse populations. However, the evidence supporting the reliability of scores produced for the re-experiencing, avoidance/numbing and arousal subscales is less strong. Significant moderating effects were found for various sample and methodological variables, such as the gender composition of the sample, cultural group, cultural orientation of the country of origin and trauma type. Building upon the findings of study 1, I performed validity generalization (VG) analyses to assess the overall construct validity of the HTQ symptom scales in Study 2. Seventy-five independent samples were included in the VG that evaluated the convergent and discriminant validity

properties of both the HTQ-16 and HTQ-30. The findings revealed that the convergent validity properties of the HTQ-16 are supported to some extent, but the discriminant validity properties are not. Furthermore, there was limited support for either the convergent or discriminant validity of the HTQ-30. Several significant moderating effects were also found for both scales (i.e. age, gender, cultural group, recruitment site, trauma type, being an original sample). Although these studies shed some light into the overall psychometric strength of the HTQ symptom scales, the decision whether to use this instrument for the assessment of PTSD should also be guided by evidence-based assessment guidelines.

Acknowledgements

I would like to express sincere gratitude to Dr. John Hunsley, my thesis supervisor, who has guided me throughout my graduate studies and provided consistent and invaluable support. Merci infiniment John. I am also indebted to my thesis committee members, Drs. Andrea Ashbaugh, Catherine Lee and Dave Miranda, and my external examiner, Dr. Paul Frewen. Your expertise and thoughtful feedback have made this thesis stronger and helped me develop as a researcher. Thank you to my friends and colleagues who shared their experiences with me and allowed me to share mine. Enfin, j'exprime ma profonde reconnaissance à ma famille qui a toujours été là pour moi. Je vous aime beaucoup.

Statement of Co-Authorship

The two manuscripts included in this dissertation were prepared in collaboration with my dissertation supervisor. I was primary author and Dr. John Hunsley was the secondary author for the first manuscript, entitled “The Harvard Trauma Questionnaire: A Reliability Generalization Study”, and the second manuscript entitled “A Validity Generalization Study of the Harvard Trauma Questionnaire.” As the primary author on all manuscripts, I was responsible for the conceptualization of the research questions and methods, coding, planning and execution of statistical analyses, and preparation of manuscripts. Dr. Hunsley was involved in coding and providing guidance and assistance in all aspects of the project, especially in the refinement of the research questions and methods, and editing of the manuscripts.

Table of contents

Abstract……...iii

Acknowledgements……..v

Statement of Co-Authorship……...……vi

List of Tables and Figures……xii

General Introduction……1

Overview of the Dissertation……..3

Psychometric Properties of Clinical Assessment Instruments……4

Reliability……...4

Internal Consistency……..7

Validity……...8

Evaluating the Psychometric Properties of Clinical Instruments……..10

Reliability Generalization……12

Validity Generalization……..……..14

Cross-Cultural Assessment……...…….15

Posttraumatic Stress Disorder (PTSD): An Overview……..……18

Prevalence of Trauma and PTSD……21

Cultural Considerations of Trauma……..23

The Assessment of PTSD……...……..27

Assessment of PTSD Across Cultures……29

The Harvard Trauma Questionnaire (HTQ)……..31

Potential Moderators of Reliability and Validity……..37

The Current Studies……..42

Study One - The Harvard Trauma Questionnaire: A Reliability Generalization Study……43

Abstract……44

Introduction……..45

The Harvard Trauma Questionnaire……46

Reliability and Reliability Generalization (RG)……..49

The Current Study……50

Method……..53

Literature Search and Data Collection……53

Inclusion Criteria……53

Data Reduction……54

Coding of Descriptive Data and Moderators……...56

Data Analysis……...58

Calculating Mean Effect Sizes……58

Random-Effects Model……59

Heterogeneity of Effects……..59

Spearman-Brown Prophecy Formula……..60

Possibility of Publication Bias……60

Analysis of Moderators……60

Results……...61

The HTQ Symptom Scales in the Empirical Literature……...61

Reporting Practices of Reliability Estimates……62

Overview of Studies Included in the Meta-Analyses and Descriptive Statistics of Potential Moderators……62

Reliability Generalization Analyses……63

Spearman-Brown Formula……...64

Publication Bias……...64

Moderator Analyses……64

Discussion……65

The HTQ in the Empirical Literature……...65

RG Analyses……66

Moderator Analyses……..68

Limitations of the Study……...71

Implications and Recommendations……72

References……74

Study Two: A Validity Generalization Study of the Harvard Trauma Questionnaire……...90

Abstract……91

Introduction……...92

The Harvard Trauma Questionnaire……92

Validity and Validity Generalization (VG)……..94

The Current Study……96

Method……..98

Literature Search and Data Collection……98

Coding of Descriptive Data and Moderators……...……99

Data Analysis……..……...101

Calculating Mean Effect Sizes……..……101

Random-Effects Model……..……102

Heterogeneity of Effects……..……..……102

Possibility of Publication Bias……..……102

Analysis of Moderators……..103

Results……103

Overview of Studies Included in the Meta-Analyses and Descriptive Statistics of Potential Moderators……..103

Validity Generalization Analyses……...105

Convergent Validity…….....105

Discriminant Validity……106

Publication Bias……106

Moderator Analyses……...107

Discussion……...108

Moderators ……112

Limitations of the Study……115

Implications and Recommendations……...116

References……...118

General Discussion……..132

The Psychometric Properties of the HTQ Symptom scales……132

The Role of Moderators……..137

Limitations of the Studies……...141

Implications……143

Recommendations for Researchers……146

Recommendations for Clinicians……148

Conclusion……..149

References (General Introduction and Discussion)…… ……152

Appendix A: Reliability and Validity Generalizations Coding Manual……..181

Appendix B: Reliability and Validity Generalizations Coding Sheet…….187

Appendix C: Studies Included in the Reliability Generalization of the HTQ (Study 1)…….192

Appendix D: Studies Included in the Validity Generalization of the HTQ (Study 2)…….207

List of Tables and Figures

General Introduction

Table 1: Criteria for Rating Psychometric Properties of Instruments……12

Table 2: Summary of Recommendations for Conducting RG……...13

Study 1

Figure 1: Literature Search Flow Chart…….85

Table 1: Descriptive Statistics of the Adaptation and Administration Procedures Reported in Samples Using the HTQ Symptom Scales ……..86

Table 2: Descriptive Statistics of Studies Included in the RG……..87

Table 3: Descriptive Statistics for the Reliability Values of the HTQ Symptom Scales and Subscales..……88

Table 4: Summary of Moderating Effects for the Internal Consistency of the HTQ Symptom Scales…….89

Study 2

Figure 1: Literature Search Flow Chart……...128

Table 1: Descriptive Statistics for the Convergent Validity of the HTQ Symptom scales……..129

Table 2: Descriptive Statistics for the Discriminant Validity of the HTQ Symptom scales…...130

Table 3: Summary of Moderating Effects for the Construct Validity of the HTQ Symptom scales……...131

General Introduction

Evidence-based practice (EBP) has become increasingly prominent in health care, including mental health services (Hunsley & Mash, 2007). Although the focus of EBP in mental health has been mainly on treatment, evidence-based assessment (EBA) has been highlighted in various EBP guidelines (e.g., American Psychological Association Presidential Task Force on Evidence-Based Practice, 2006; Canadian Psychological Association Task Force on Evidence-Based Practice of Psychological Treatments, 2012). A key principle underlying EBA is that research and theory should be used to guide decision-making regarding assessment targets, methods and measures, and the assessment process itself (Hunsley & Mash, 2007). This can lead to more consistency regarding diagnosis across clinicians, researchers, and settings, and also to the provision of higher quality treatment (Speroff et al., 2012). An important element of EBA is using psychometrically strong measures, that is, ensuring that the reliability and validity properties of an instrument are sound when used with a specific sample for a specific purpose (Hunsley & Mash, 2007). For example, if a clinician is planning to use an instrument in the process of determining whether a Vietnamese refugee meets criteria for a diagnosis of PTSD, the clinician must ensure that the instrument he or she decides to use has supporting psychometric evidence for diagnostic purposes within a Vietnamese refugee population.

Psychodiagnostic assessments rely heavily on the conceptualization of the construct they are intended to measure. If the conceptual underpinnings of the construct were not solid, then the psychometric properties of the instrument would be compromised (Haynes, Smith, & Hunsley, 2011). One psychological construct that has been heavily debated in regards to its conceptualization is the posttraumatic stress disorder (PTSD) diagnosis based on the various editions of the American Psychiatric Association (APA)'s Diagnostic and Statistical Manual of

Mental Disorders (DSM). A central argument is that the PTSD construct was developed according to Western cultures' conceptualizations of trauma and its sequelae, and may not be applicable to other cultures (e.g., Jones et al., 2003). Consequently, it is not surprising that there has also been much questioning regarding the cross-cultural applicability of screening tools used to assess PTSD. Although research in cross-cultural assessment in general has received increased attention and several authors have proposed guidelines for validating instruments for different cultures (e.g., International Test Commission, 2005), the assessment of PTSD across cultures is especially challenging due to variations in trauma symptom expression (e.g., Pole, Gone, & Kulkarni, 2008).

A trauma assessment instrument specifically developed to be used and adapted across cultures is the Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992). Systematic reviews of tools used to assess the health of refugees have found that the HTQ has been quite extensively used by researchers to assess trauma and its sequelae (Gagnon, Tuck, & Barkun, 2004; Hollifield et al., 2002). Among the various PTSD instruments evaluated in these reviews, the HTQ was described as being a scientifically strong measure because of the procedures used in its development and its psychometric properties. In addition, the HTQ has been recommended by other experts in the field of trauma assessment such as Keane, Silberbogen, and Weierich (2008) and Nakeyar and Frewen (2016). Although this instrument has been recommended above others to assess PTSD in different cultures, a more detailed investigation of individual studies that have used the HTQ suggests that its psychometric properties may not be so robust across populations (e.g., Jakobsen, Thoresen, & Johansen, 2011; Rasmussen, Verkuilen, Ho, & Fan, 2015; Silove et al., 2007). Given the popularity of this instrument and the inconsistent findings among the few

validation studies available on the HTQ, a further investigation into the overall psychometric strength of the HTQ is needed.

Overview of the Dissertation

This dissertation contributes to the EBA literature by examining the empirical evidence of the psychometric properties of the HTQ symptom scales. The goal of this research is to provide information regarding the cross-cultural applicability of the HTQ symptom scales, thereby helping clinicians and researchers make evidence-based decisions regarding their use of this instrument with individuals from different cultures. The data for two meta-analytic studies were extracted from empirical publications and dissertations that used the HTQ symptom scales in their own data collection. These studies were coded and the information obtained was used to (a) provide descriptive statistics of how the HTQ symptom scales were used in the empirical literature and (b) evaluate its psychometric properties, specifically reliability and validity. The first study included a reliability generalization (RG) study in which I calculated overall mean reliability estimates (i.e., Cronbach's alpha) across studies. The second study is a validity generalization (VG) study in which I examined the evidence for convergent and discriminant validity properties of the HTQ symptom scales across the identified primary studies. I conducted moderator analyses in both studies to assess whether sample and/or methodological characteristics had a significant impact on the mean estimates of these psychometric properties.

The general introduction of this dissertation is comprised of six sections. The first is an overview of the psychometric properties of clinical assessment instruments, which is then followed by a discussion on the psychometric evaluation of clinical instruments. In the third section, I discuss important features of cross-cultural assessment. This section is followed by an overview of PTSD, which includes the prevalence of the disorder and of traumatic events,

possible cultural elements of PTSD, and a description of the assessment of PTSD, both in general and across cultures. Next, I provide a description of the development and psychometric properties of the Harvard Trauma Questionnaire. In the final section, I present the variables to be examined as potential moderators of the reliability and validity of the HTQ.

Psychometric Properties of Clinical Assessment Instruments

Measurement is an inherent part of psychological research and clinical work. Researchers and clinicians regularly make important decisions in the selection of appropriate assessment strategies as they approach a new case or develop a new research protocol. This entails selecting the most appropriate measurement instruments, or more precisely, instruments that are scientifically supported for both the purpose and population at hand. Thus, understanding the properties that affect the scientific quality of psychological measures is crucial throughout this decision-making process. Two fundamental concepts that compose the scientific foundations of measurement are reliability and validity. These concepts are described in the following sections.

Reliability. Reliability, or the precision of a measure, represents "the consistency of the scores across instances of the testing procedures" (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p. 33). Furthermore, reliability estimates represent the proportion of variance that is explained by the true score itself (characteristics measured in the test), as opposed to the error score, the part of the variance that is explained by sampling or measurement error (Graham, Yenling, & Jeziorski, 2006). As the reliability of test scores is also a prerequisite for the validity of an instrument, examining the degree of consistency of scores is imperative in determining if the instrument is indeed scientifically sound.

There are four main types of reliability: (a) test-retest, (b) interrater, (c) parallel forms, and (d) internal consistency (Salkind, 2012). Test-retest reliability refers to temporal stability, and can be operationalized as a correlation of scores of a test that has been administered at different time-points (Haynes et al., 2011). Interrater reliability can be defined as the assessment of the agreement between two or multiple raters. There are different types of interrater reliability, including percentage agreement and interrater correlations (Haynes et al., 2011). Parallel forms reliability refers to the examination of different forms of the same instrument (Salkind, 2012). Finally, internal consistency refers to the degree of consistency among the scores on items of an instrument (Haynes et al., 2011) and can be evaluated by comparing two halves of an instrument (split-half reliability), or by computing an index such as Cronbach's alpha (Cronbach, 1951). Cronbach's alpha is the most commonly used and reported index of reliability (Tavakol & Dennick, 2011) and it is often used in meta-analytic studies of the reliability properties of an instrument (Vacha-Hasse & Thompson, 2011). Details regarding internal consistency and Cronbach's alpha are described further.

When describing reliability, it is also important to note that it refers to the properties of test scores of a specific sample and not to the properties of an instrument itself. Saying "this test is reliable" rather than "this test produces reliable scores for this sample" is one of the most common errors made in psychology with regard to measurement (Thompson, 1994). Indeed, factors such as sample characteristics can dramatically influence the degree of consistency of test scores. For instance, the same instrument can produce scores with differing reliability when administered to more or less homogenous samples (Thompson, 1994). In sum, reliability is not only dependent on the quality of a measure, but also the characteristics of the sample, sample size, and administration and scoring procedures used (Barnes, Harp, & Jung, 2002). Therefore,

gathering reliability evidence is especially important for measures that are commonly used with various types of samples, such as the HTQ.

Researchers often report reliability estimates of prior studies but do not report the data from their own studies (Vacha-Hasse, Kogan, & Thompson, 2000). Ignoring the sample-specific nature of reliability may then lead researchers to interpret unreliable data (Henson, Kogan, & Vacha-Haase, 2001). Although the American Psychological Association Task Force on Statistical Inference specified that all studies should include the reliability coefficients of the scores of their sample even if the study is not psychometric in nature (Wilkinson & American Psychological Association Task Force, 1999), researchers continue to omit reliability coefficients from reports of their studies. To examine this widespread tendency, Vacha-Haase and Thompson (2011) conducted a review of 47 previous meta-analytic studies of reliability. Their results showed that 54% of the 12,994 primary studies included in their review made no mention of reliability in their studies, whereas 15.7% mentioned reliability but only provided the coefficients reported in previous studies.

This tendency to generalize reliability properties across studies has been described by Vacha-Haase et al. (2000) as "reliability induction." They argued that reliability induction is only plausible if the previous samples have sample composition and score variability that are similar to the current set of participants. Even when samples may seem to be comparable, examining score reliability is still important as there may be subtle differences between samples. One way of examining the generalizability of the reliability of test scores is by conducting a reliability generalization (RG) study, which is a meta-analytic method described in further detail in a subsequent section.

Internal consistency. Internal consistency refers to the degree to which all the items in a test measure the same construct. Cronbach's alpha is a measure of the inter-relatedness of test items and has a value between 0 and 1. If the items of a test are highly correlated to each other, Cronbach's alpha will be higher (Tavakol & Dennick, 2011). Calculating Cronbach's alpha is particularly useful when examining the reliability properties of an instrument that does not have right or wrong answers (Salkind, 2012), such as the HTQ-symptom scale. It is important to note, however, that Cronbach's alpha is affected by the number of items on a test. Therefore, there can be high internal consistency for other reasons, such as redundant items or using a large number of items (Haynes et al., 2011). Other difficulties and criticisms associated with coefficient alpha include: (a) it is often based on unmet assumptions (e.g. assuming that the true score variance is the same across all items); (b) these assumptions can inflate alpha, thus providing imprecise estimations of internal consistency; (c) alpha cannot be generalizable if an item is deleted, and (d) variability is not accounted for in a point estimate of alpha (e.g. Dunn, Baguley, & Brunsden, 2014). Some strategies have been developed to address these concerns about alpha. For instance, researchers can use bootstrapping methods to produce confidence intervals around the alpha point estimate to provide a range of probable values (Dunn et al., 2014). Preferably, researchers are recommended to calculate omega (McDonald, 1999), a different measure of internal consistency, which has shown to be a more accurate index as compared to alpha (e.g. Zinbarg, Revelle, Yovel, & Li, 2005). However, until this alternative becomes more accessible and used in the literature, Cronbach's alpha currently remains the best available measure of internal consistency that can be examined by meta-analysis.

Establishing internal consistency is important as it can affect the accuracy of interpretation of scores and is relevant in both clinical and applied research settings. Calculating

typical reliability estimates help determine the effect of measurement error on the observed score of a sample of participants in research settings (Haynes et al., 2011). In clinical settings, a typical internal consistency value can also be used to reveal the effect of measurement error on the observed score of an individual through the calculation of the standard error of measurement (Haynes et al., 2011; Tavakol & Dennick, 2011). The standard error of measurement is then used to calculate confidence intervals, providing a range of values that is likely to contain an individual's true score. In addition, producing mean internal consistency estimates for samples allows clinicians and researchers to calculate a reliable change index, which measures the extent to which an individual makes gains in treatment (Haynes et al., 2011).

As described earlier, reliability is an important step in establishing validity. In other words, an instrument needs to produce reliable scores in order to produce valid scores. The internal consistency of an instrument can be used to determine the maximum value of the validity coefficient that can be obtained. The index of reliability, or the square root of the reliability estimate, represents the highest validity estimate of the scores produced by the same sample (Haynes et al., 2011). Once the degree of consistency of the scores of an instrument is assessed, the next step in the evaluation of a measure is to examine the evidence of validity. The index of reliability can thus serve as an evaluation benchmark of the validity coefficients.

Validity. Validity refers to the extent to which an assessment instrument measures what it is supposed to measure. In other words, an instrument that has been evaluated as having good evidence of validity has variation in scores that reflects the variation in the construct that is to be measured (Haynes et al., 2011). According to the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014, p. 11), validity can also be defined as "the degree to which

evidence and theory support the interpretations of test scores entailed by proposed uses of tests" and is considered the most fundamental element of test development and evaluation.

Similar to reliability, there are various kinds of evidence of validity. The four main categories are as follows: (a) content validity, (b) criterion validity, (c) incremental validity, and (c) construct validity. Content validity refers to the extent to which items of an instrument are representative of the construct that is intended to be measured (Haynes, Richard, & Kubany, 1995), or that the test items cover the range of elements needed to assess a certain construct. Criterion validity relates to whether an instrument is associated with other current or future outcomes that are measured by different instruments, and is often described as either predictive or concurrent: predictive validity is when a test score predicts a certain outcome, and concurrent validity is when a test score is related to an outcome that is measured simultaneously (Haynes et al., 2011). Incremental validity entails the question of whether the information retrieved from one or more measures increases the validity or utility of a clinical decision more than other already established sources of information (Hunsley & Meyer, 2003). Finally, construct validity refers to the degree to which a test score reflects the construct of interest, and includes both convergent and discriminant validity. Convergent validity refers to the extent to which the scores of an instrument are related to scores of a different measure that is intended to assess the same construct, or other variables that have theoretically established associations with the construct (Messick, 1995). Discriminant validity refers to the degree to which the scores of an instrument are unrelated to measures of unrelated constructs (Haynes et al., 2011). Thus, low correlations between the two instruments can be indicative of good discriminant validity.

Validity is not a static characteristic of an instrument but can vary according to context and different sample characteristics. This was described as "situational specificity" by Schmidt

and Hunter (1998) and was used to support their hypothesis that there are statistical artifacts (e.g., sampling error, measurement error) that can cause instruments to demonstrate adequate validity in one situation and inadequate validity in another. This echoes the concept that reliability estimates are sample-dependent. Those using an instrument are encouraged to examine validity evidence for their specific population and purpose. This evidence for validity can be collected through various means (AERA, APA, & NCME, 2014) including: (a) test content (e.g., content items reflect the construct of interest), (b) response processes of test-takers (e.g., looking at how respondents interpret items), (c) internal structure of the instrument (e.g., examining the relations among items through factor analysis), (d) consequences of completing the instrument, and (e) relations to other variables. This last method of examining validity properties includes convergent and discriminant evidence, test-criterion relations, and validity generalization, which is described in more detail in a subsequent section.

Establishing the validity of scores produced by an instrument is important as it also speaks to the validity of theories on the construct being measured and the operational definition of the construct in question (Haynes et al., 2011; Kimberlin & Winterstein, 2008). The evidence of validity of a measure continuously evolves as the theories of the constructs being measured change according to new research developments (Haynes et al., 2011). Thus, evaluating the validity of measures should be an ongoing process. This most certainly applies to the HTQ, given the changes in the PTSD construct in the most recent version of the DSM and the cultural variation of symptoms.

Evaluating the Psychometric Properties of Clinical Instruments

As highlighted, establishing psychometric evidence is an inherent component of the development and adaptation of assessment instruments. Once the reliability and validity

estimates have been calculated, it is then important to determine whether or not the values obtained are adequate for using and interpreting the data provided by an instrument. Consequently, several experts have proposed criteria to evaluate assessment tools (e.g., Bickman et al., 1999; Hunsley & Mash, 2008; Robinson, Shaver, & Wrightsman, 1991) to help clinicians and researchers determine the suitability of measures for a specific purpose.

Hunsley and Mash's (2008) criteria can be applied to measures that assess mental health issues across populations. They developed a rating system that includes nine psychometric categories: norms, internal consistency, inter-rater reliability, test-retest reliability, content validity, construct validity, validity generalization, sensitivity to treatment change, and clinical utility. For each category, the following ratings can be assigned: less than adequate, adequate, good, excellent, unavailable, and not applicable. A rating of 'adequate' indicates that an instrument attains a minimal level of scientific support, 'good' indicates that there is solid scientific evidence that supports the instrument, and the rating of 'excellent' indicates extensive, high quality evidence supporting the instrument. The authors recommended, for both research and clinical purposes, only using instruments that fall in the 'good' or 'excellent' ranges. See Table 1 for a description of ratings for select psychometric categories. Because the focus of this dissertation is on internal consistency, convergent validity, and discriminant validity, only the categories of internal consistency and construct validity are presented. Hunsley and Mash's rating system for these categories also served as guidelines when interpreting the results of the two dissertation studies.

Table 1

Criteria for Rating Psychometric Properties of Instruments

Internal consistency
Adequate: Preponderance of evidence indicates alpha values of .70-.79.
Good: Preponderance of evidence indicates alpha values of .80-.89.
Excellent: Preponderance of evidence indicates alpha values of ≥ .90.
Construct validity
Adequate: Some independently replicated evidence of construct validity (e.g., predictive validity, concurrent validity, convergent and discriminant validity).
Good: Preponderance of independently replicated evidence, across multiple types of validity (e.g., predictive validity, concurrent validity, convergent and discriminant validity).
Excellent: In addition to the criteria used for a good rating, evidence of incremental validity with respect to other clinical data.

Note. Adapted from "Developing Criteria for Evidence-Based Assessment: An Introduction to Assessments That Work," by J. Hunsley and E. J. Mash, 2008, *A Guide to Assessments That Work*, pp. 8-9.

Reliability generalization. As reliability estimates vary across samples and different administrations of a given instrument, it can be difficult to assess its overall reliability properties. To obtain comprehensive information on reliability of scores for a given instrument, researchers conduct meta-analyses that combine data from studies that have used the instrument (e.g., Churchill & Peter, 1984; Peter & Churchill, 1986). There has been a marked increase in interest and the practice of conducting meta-analyses on reliability coefficients since Vacha-Haase (1998) coined the term "reliability generalization" (RG). In general, RG studies are designed to: (a) determine a mean reliability of scores generated by a test, (b) examine the variability in score reliability, and (c) characterize the sources of variance by examining the relations between the reliability estimates and study and sample characteristics. This method is an extension to the "validity generalization" procedures developed by Schmidt and Hunter (1977) that examines the relations between scores from a given measurement instrument and other criteria as a means to evaluate the overall validity properties of the instrument.

RG vary considerably in their methodology (López-López, Botella, Sánchez-Meca, & Marín-Martínez, 2013). In her dissertation, Henchy (2012) summarized RG best practice guidelines found in the existing literature (e.g., Howell & Shields, 2008; Thompson, 1999; Warne, 2008). These guidelines are presented in Table 2. Essential recommendations include important RG practices that have been commonly found and recommended in previous research, whereas optimal recommendations are somewhat newer practices that should be considered. Thus, these guidelines can help researchers conduct RG that are consistent with best practices. They served as a roadmap for the present RG and, with the exception of the use of a power analysis (because of the availability of scores of studies), are included in the methodology of this dissertation.

Table 2

Summary of Recommendations for Conducting RG

Essential recommendations
Conduct thorough searches for studies
Determine criteria for inclusion of primary studies
Address file-drawer problem or publication bias
Code both instrument and sample characteristics
Use multiple raters and report inter-rater reliability
Use box and whisker plots or Confidence Intervals to present variability
Optimal recommendations
Do not combine multiple types of reliability
Do not combine multiple subgroups in one analysis
Conduct separate analyses for multiple subscales or instruments
Examine homogeneity of population correlations
Conduct a power analysis before conducting the RG

Note. Adapted from "Review and Evaluation of Reliability Generalization Research", by A. M. Henchy, 2012, *Unpublished Doctoral Dissertation*, pp. 34-36.

RG analyses for the field of psychology can help those who use a given test make decisions regarding the applicability of the test to a specific group in research settings or to individuals in clinical settings (Vacha-Haase et al., 2000). Another potential benefit of

conducting RG, especially in multicultural research, is its potential effect on public policy and the development of treatment programs for ethnic minorities (Chun, Organista, & Marín, 2003). For instance, if an instrument that is being used to assess posttraumatic symptoms in a given ethno-cultural group has not yet demonstrated adequate reliability properties for this specific group, the results may be misinterpreted and could have a negative impact on the decision-making process regarding access to services. Values generated by RG analyses may provide an indication of the range of reliability estimates that are applicable for different groups. Thus the importance of the evaluation of psychometric properties of instruments cannot be underestimated, and more awareness of the potential impacts of instruments with inadequate or unknown psychometric information needs to be considered.

Validity generalization. Following their initial observations regarding situation specificity in validity, Schmidt and Hunter (1977) developed a meta-analytic method called "validity generalization" (VG) to examine the correlations between a specific test and a criterion. Specifically, VG is intended to estimate a mean validity coefficient of a measure based on the correlations between the measure and validity criteria of different studies. For example, to evaluate the discriminant validity properties of a PTSD questionnaire, one could gather relevant studies that provide correlations (or other measures of the relations between two variables) between this questionnaire and a measure of impression management, and then calculate a mean correlation. This method can also provide indications of the effects that might account for variability in the values found in a given sample, and is similar to the RG meta-analytic procedures proposed by Vacha-Haase (1998).

An important issue that needs to be mentioned is the assessment of the strength of the relation between two variables. In other words, how can one determine if a correlation is strong

or weak? Although *r*-values range from -1 to +1 (-1 = perfect inverse correlation, 0 = no correlation, +1 = perfect positive correlation), it is difficult to establish the strength of values across the full range of possible values. However, Cohen (1992) proposed the following *r*-value effect-size benchmarks to help with interpreting correlations: .10 = small effect, .30 = medium effect, .50 =large effect. Although these benchmarks have been widely used in various fields of research, Hemphill (2003) suggested other benchmarks based on 380 meta-analytic studies in applied psychological research that are more appropriate for psychology (i.e. <.20 = small effect, .20-.30 = medium effect, >.30 = large effect).

The VG method has been primarily used in the field of employment selection and testing (e.g., Dye, Reck, & McDaniel, 1993; Ones, Viswesvaran, & Schmidt, 2003) and it has been recommended in the Standards for Educational and Psychological Testing as a general approach to evaluate validity (AERA et al., 2014). According to these standards, VG should preferably be used when the following conditions are met: a large meta-analytic database of data representing more or less the type of situation to which the use of scores on the instrument will be generalized, and when statistical artefacts are corrected and this correction produces validity evidence that is consistent (AERA et al., 2014).

Cross-Cultural Assessment

A central aspect of cross-cultural assessment is the question of whether test scores can be interpreted the same way across various cultural groups. Thus, in order for a psychological assessment instrument to be used among different cultures and sub-groups, many elements need to be considered. Two concepts that are critical to this issue are bias and equivalence (e.g. He & van de Vijver, 2012).

Bias refers to various elements that can impact the accuracy and validity of instruments across cultures and it occurs when score differences between cultural groups do not reflect the actual cultural differences of the underlying construct being measured (van de Vijver & Tanzer, 2004). As He and van de Vijver (2012) summarized in their review, the several types of bias can be classified as being either construct bias (i.e. a construct being measured is different across cultural groups), method bias (i.e. bias deriving from the sampling procedures, structure of an instrument, different response styles or the administration process), or item bias (i.e. an item that has a different psychological meaning across cultures).

Measurement equivalence implies that a same instrument measures the same construct across diverse cultural groups. There are several types of equivalence, including: (a) content equivalence; (b) linguistic and semantic equivalence; (c) conceptual equivalence; (d) scale and technical equivalence; (e) normative equivalence (Keane et al., 1996). Content equivalence refers to the notion that the items of a measure reflect the ways that individuals of a culture experience the phenomenon that is being measured. Using the psychological effects of a traumatic event as an example, if several items measure the presence of numbing symptoms, but numbing symptoms are not a typical posttraumatic reaction within the culture in question, the instrument would be lacking in content equivalence. Linguistic and semantic equivalence ensures that the grammar, syntax, and meanings of a translated instrument are comparable to the original version. The most commonly recommended method to ensure this type of equivalence is to use both a blind-back translation and a consensus approach to determining the phrasing of items (Brislin, 1970, 1986). Conceptual equivalence refers to the idea that a concept is equivalent across cultures. More specifically, it addresses the question of whether a translated expression assesses the same aspect or construct in different cultures. Scale and technical equivalence ensures that

the methods used to assess a construct are comparable across cultures and yield scores on an instrument that mean the same thing across groups. Lastly, normative equivalence pertains to the establishment of normative standards specific to the culture in question, instead of relying on pre-established norms developed in a different culture.

The assessment of the quality of adapted instruments must take into account the methods that were used to ensure these types of equivalencies, as well as verify the psychometric properties of the instrument. A methodological review of 47 instrument translation studies revealed that there was great variability in the methods used to translate and validate instruments used for cross-cultural research (Maneesriwongul & Dixon, 2004). In addition, Sperber (2004) noted that the process of translating and cross-culturally validating instruments is usually not considered important in research protocols and that the most common translation procedure used is the simple forward translation.

Many different guidelines and recommendations for the cross-cultural adaptation of instruments exist, however very few of them have been empirically investigated. A cross-disciplinary review of cross-cultural adaptation guidelines identified 31 of them and found no consensus in their methodology (Epstein, Santo, & Guillemin, 2015). However, this review showed that most guidelines included the recommendation of using committees, focus groups, and back translations. Although the empirical evidence regarding these guidelines is lacking, it appears that the use of back translations may not have much added value if a consensus approach is already used (e.g. da Mota Falcão, Ciconelli, & Ferraz, 2003; Epstein, Osborne, Elsworth, Beaton, & Guillemin, 2013). Although currently there is limited empirical support for these guidelines, using strategies to help minimize bias remains unquestionably essential in cross-

cultural measurement. If the adaptation process is not carefully implemented, the results may be invalid (Epstein et al., 2015).

Posttraumatic Stress Disorder: An Overview

The psychological responses after experiencing traumatic life events have been of interest for many centuries. For instance, authors such as Homer and Shakespeare wrote about combat stress that appears to closely resemble current PTSD symptomatology (Trimble, 1981). Following these early depictions, such post-trauma reactions became more recognized as syndromes in the 19th and 20th centuries, with a diverse nomenclature being used, including spinal concussion, soldier's heart, traumatic neurosis, and shell shock, to name a few (Parry-Jones & Parry-Jones, 1994).

Although early editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM) included disorders that occurred as a consequence of a stressful event such as "gross stress reactions" (DSM-I: APA, 1952), and "anxiety neurosis/transient situational disturbance" (DSM-II: APA, 1968), PTSD was first recognized as a distinct diagnosis in the third edition of the DSM (DSM-III: APA, 1980). Since then, the PTSD diagnosis has undergone many changes throughout the subsequent DSM editions, with most modifications involving the stressor criterion, specific symptoms and specifiers. In the DSM-5 (APA, 2013), the PTSD diagnosis was moved from the "Anxiety Disorders" chapter to a newly created one entitled "Trauma and Stress-Related Disorders." The stressor criterion was significantly changed (e.g., Criterion A_2 that required that the individual must have reacted to the traumatic event with "fear, helplessness, or horror" (APA, 2000, p. 467) was removed, and dissociative and preschool subtypes were added. Because the HTQ symptom scales are based on the 17 symptoms of Criteria B, C and D

of the DSM-IV (APA, 1994) PTSD diagnosis, I will specifically focus on these symptoms and describe them in light of the changes in the current DSM-5 (APA, 2013).

One major change to the PTSD symptoms in the DSM-5 was a new conceptualization of the factor structure of the disorder. The DSM-IV edition described a three-factor model, or three symptom clusters: (a) Criterion B: re-experiencing, (b) Criterion C: avoidance/numbing, and (c) Criterion D: hyperarousal (APA, 1994). However, numerous confirmatory factor analyses revealed that this three-factor model was not the best to capture the underlying symptoms structure of PTSD (e.g., Baschnagel, O'Connor, Colder, & Hawk, 2005; Elklit & Shevlin, 2007; Krause, Kaltman, Goodman, & Dutton, 2007). Although a range of one to seven-factor models have been proposed, the model that has received the most support in the literature is a four-factor conceptualization, and the APA sub-group included this one in the DSM-5 (Friedman, 2013). This model is primarily based on King, Leskin, King, and Weathers' (1998) conceptualization, which separates avoidance and numbing into two clusters. The DSM-5 model includes the following symptom clusters: (a) Criterion B: intrusion (e.g., "recurrent, involuntary, and intrusive distressing memories of the traumatic event," APA, 2013, p. 271), (b) Criterion C: avoidance (e.g., "avoidance of or efforts to avoid distressing memories, thoughts, or feelings about or closely associated with the traumatic event(s)," APA, 2013, p. 271, (c) Criterion D: negative alterations in cognitions and mood (e.g., "feelings of detachment or estrangement from others," APA, 2013, p. 272) and (d) Criterion E: alterations in arousal and reactivity (e.g., "exaggerated startle response," APA, 2013, p. 272). More recently, a seven-factor hybrid model that includes re-experiencing, avoidance, negative affect, anhedonia, externalizing behaviours, and anxious and dysphoric arousal symptom clusters has preliminary evidence of being a superior fit as compared to all other models (Armour, Müllerová, & Elhai, 2016).

The DSM-5 edition contains all the 17 symptoms of DSM-IV, however some descriptions have been revised and reworded to provide further clarification. Three new symptoms were also added. One example of an added item is "reckless or self-destructive behaviour" (APA, 2013, p. 272) in the Criterion E section. This is intended to reflect the fact that externalizing behaviours such as risk-taking and suicidal behaviours are common in individuals with PTSD (e.g., Miller, Kaloupek, Dillon, & Keane, 2004; Nock, Hwang, Sampson, & Kessler, 2010). This increase in the number of symptoms to 20 expanded the possible criteria to meet the diagnosis and, consequently, increased the number of possible symptom profiles. In addition, although the diagnostic thresholds were not changed (i.e., a minimum of six symptoms is still needed to meet criteria), they have been redistributed among the new symptom clusters. In other words, instead of requiring one re-experiencing symptom, three avoidance/numbing symptoms, and two arousal symptoms to meet full PTSD criteria, an individual must experience one intrusion symptom, one avoidance symptom, two negative alterations in cognitions and mood symptoms, and two alterations in arousal and reactivity symptoms.

Studies that have compared the DSM-IV and DSM-5 PTSD diagnostic criteria have found differences in prevalence rates between the two (e.g., Calhoun et al., 2012; Gentes et al., 2014; Kilpatrick et al., 2013). A study conducted of a community sample in Wales found that the DSM-IV prevalence was 14.3%, but only 8% when DSM-5 criteria were applied (White et al., 2015). The change in prevalence was attributed to the exclusion of DSM-IV stressor criterion qualifying events such as life-threatening illnesses for DSM-5. Changes in diagnostic rates in other studies were also attributed to the splitting of avoidance and numbing symptoms into two distinct clusters (e.g., Gentes et al., 2014).

Studies on the rate of PTSD across different cultures using DSM-5 criteria are limited

due to the recency of its publication. Therefore, it is too early to establish the cross-cultural applicability of the new criteria. However, considerable data regarding the prevalence rates across world have been gathered with the past versions of DSM. These studies are summarized in the following section.

Prevalence of trauma and PTSD. It is important to note that the methods used to collect data, the diagnostic criteria used to evaluate trauma and PTSD (e.g., DSM-IV vs. DSM-5), and the composition of research samples (e.g., war veterans vs. civilians; refugee vs. non-refugee) can all have a tremendous impact on the rates of trauma and PTSD reported in the literature. Thus, it is critical that these factors are kept in mind when interpreting results of epidemiological studies. There are also three different concepts to be aware of when studying the epidemiology of trauma: (a) the prevalence of being exposed to traumatic events, (b) the conditional prevalence, or the risk of having PTSD after being exposed to a traumatic event, and (c) the total prevalence of PTSD in the population (Norris & Slone, 2014).

With regards to the lifetime prevalence of trauma exposure, studies have shown that experiencing traumatic events is quite common among general populations across the world. For example, Kilpatrick and colleagues (2013) found in a national probability study of approximately 3,000 U.S. adults that 89% of participants reported being exposed to at least one DSM-5 qualifying event. One study on the exposure to trauma in four post-conflict, low income countries/regions (i.e., Algeria, Cambodia, Ethiopia, and Gaza) found that the rates of trauma according to various trauma categories (e.g., torture, youth domestic stress, and conflict-related events) varied across countries, but that overall exposure to severe trauma was consistently high (de Jong et al., 2001). The most recent data collected from the World Mental Health Surveys that

were conducted in 24 countries between 2001 and 2012 show rates of lifetime trauma exposure ranging from 28.6% in Bulgaria to 84.6% in Ukraine (Koenen et al., 2017).

Even though traumatic events are common, the majority of individuals who experience them do not develop PTSD. For instance, in one study based on DSM-IV criteria, 9.5% of individuals exposed to trauma developed PTSD (Breslau et al., 1998). In addition, in Kilpatrick and colleagues' (2013) study, the conditional probability of lifetime PTSD was found to be only 11.7%. It is interesting to note that although, in general, men are more likely to experience trauma than are women, women have a higher tendency to develop PTSD than do men (e.g., Breslau et al., 1998; Kessler et al., 1995). Another factor that influences the conditional probability of PTSD is type of trauma experienced. For instance, survivors of violent and intentional traumas are more likely to develop PTSD than are those who experience unintentional traumas such as natural disaster or those who are witnesses to a traumatic event (e.g., Kessler et al., 1995).

Although traumatic events and negative psychological impacts due to trauma are prevalent in all cultures (Keane et al., 2008), epidemiologists have found varying rates of the overall lifetime prevalence of PTSD across countries. Studies in the United States have shown that PTSD is present in 6.4% to 9.5% of the general population (Breslau, Davis, Andreski, & Peterson, 1991; Kessler et al., 1995; Pietrzak, Goldstein, Southwick, & Grant, 2011). The World Mental Health Surveys from 24 countries found estimates for lifetime prevalence for PTSD that ranged from 0.3% in China to 8.8% in Northern Ireland (Koenen et al., 2017). In addition, a systematic review on the prevalence of mental disorders in refugees who were resettled in Western countries showed that 9% of them suffered from PTSD (Fazel, Wheeler, & Danesh, 2005). Furthermore, research in areas of conflict or where other mass traumatic events took

place, such as natural disasters or terrorist attacks, reveals higher rates. For instance, approximately 20 months after the December 2004 Indian Ocean tsunami, the prevalence of PTSD among victims in Sri Lanka was 21% (Hollifield et al., 2008). Similarly, 19.5% of a sample of rescue and recovery workers in the aftermath of 9/11 continued to screen positive for PTSD five years after the attacks (Brackbill et al., 2009). Although there are differences in rates of PTSD among studies, it is apparent that the negative psychological effects of traumatic events are considerable among people across cultures.

Cultural considerations of trauma. The relation between culture and mental health is typically seen as one that is highly intertwined (e.g., Sam & Moreira, 2002). Cultural factors that can influence how mental disorders are experienced and expressed include ethnicity, race, acculturation, individualism-collectivism, and social acceptance of expressed distress, among others (Eshun & Gurung, 2009; Yeomans & Foreman, 2009). The degree to which culture plays a role in psychodiagnostic classification is central to the debate between emic and etic approaches to psychopathology. The emic approach considers psychopathology to be specific to individual cultures, as opposed to the etic approach which views psychopathology as universal with some sociocultural factors influencing symptom expression (Murphy, 1982).

Although most trauma research has been conducted in Western countries, more and more research has been done with people from non-Western cultures and, as the cross-cultural research on PTSD develops, more and more questions arise regarding the cross-cultural applicability of the diagnosis. Specifically, questions about face validity and clinical utility have provoked much debate (Hinton & Lewis-Fernández, 2011; Summerfield, 1999). Some researchers have argued that some PTSD criteria, such as flashbacks, are a Western phenomenon (Jones et al., 2003). Other researchers have gone further, suggesting that the psychological impact of trauma varies so

much, and depends on so many factors, that it cannot be standardized into a single disorder or diagnosis such as PTSD (e.g., Bracken, Giller, & Summerfield, 1995; Summerfield, 2004) and that local idioms of distress or cultural syndromes are better to describe the specifics of trauma response. Rather than adopt an either-or stance, a number of researchers have argued that the use of both the PTSD construct and local idioms of distress are relevant for understanding the sequelae of trauma (e.g., Miller et al., 2009).

As indicated, there has been evidence of PTSD in numerous cultures. To further illustrate, studies that have used the same instrument (i.e., Composite International Diagnostic Interview for DSM-IV-TR [CIDI]; Robins et al., 1988) to assess the presence of PTSD, have found evidence of the disorder in various countries. For example, the 12-month prevalence rate of PTSD using the CIDI was 0.4% in Japan (Kawakami et al., 2005), 0.6% in Mexico (Medina-Mora et al., 2005), 0.6% in South Africa (Williams et al., 2008) and 3.5% in the U.S. (Kessler et al., 1995). Although these studies show that the DSM PTSD construct can be found across cultures, there are significant differences in the prevalence rates reported. Numerous factors may contribute to these variations, such as the prevalence of trauma exposure and methodological variables (e.g., lack of equivalency among the various adaptations of the instrument), to name a few. However, some studies have attempted to control for these factors and have found true differences in PTSD prevalence rates. For example, Alcántara and colleagues (2013) found differences after controlling for prevalence of exposure, and de Jong and colleagues (2001) found differences after ensuring measurement equivalency.

Another factor that may contribute to the variation in the epidemiology of PTSD is the variation across cultures in the manifestation of trauma symptoms. For instance, it appears that the salience of avoidance/numbing symptoms may vary across cultures. A study of Kalahari

Bushmen found that several avoidance/numbing symptoms were not endorsed among the group of traumatized individuals, but that the re-experiencing and arousal symptom were (McCall & Resick, 2003). Other studies among Vietnamese refugees (Norris, Van Landingham, & Vu, 2009), Cambodians (Mollica, Mcinness, Poole, & Tor, 1998), Middle Eastern women (Norris & Aroian, 2008), and Senegalese refugees (Tang & Fox, 2001) found that avoidance and numbing symptoms were less salient than symptoms of the other clusters. Symptom variations have also been found with the re-experiencing cluster, with the presence of distressing dreams or nightmares being more salient in certain cultural groups, such as American Indians (Shore, Orton, & Manson, 2009), Alaskan Natives (Palinkas, Petterson, Russell, & Downs, 2004), and Cambodian refugees (Hinton et al., 2009).

Another example of culturally bound symptom variation is the higher prevalence rate in the United States of PTSD among Latinos, African Americans, and Native Americans, compared to Whites. Apart from some variations in prevalence of traumatic exposure (except for Latinos), other factors that have been hypothesized to contribute to these differences are an over-endorsement of PTSD symptoms related to cultural differences in response styles, the effects of racism and discrimination, possible overlapping symptoms with other cultural idioms of distress, and a tendency to experience peri-traumatic dissociation (Alcántara et al., 2013; Pole et al., 2008).

In addition, somatic symptoms have been found to be common reactions to trauma in certain cultures. For instance, Salvadoran and Senegalese refugees have reported sensations of bodily heat (Jenkins & Valiente, 1994; Tang & Fox, 2001), Bhutanese refugees who were tortured reported experiencing bodily heat, and several studies have also shown various somatic symptoms among Cambodian refugees such as neck pain and orthostatic dizziness (Hinton et al.,

2006; Hinton et al., 2008). Therefore, such missing symptoms from the PTSD construct may also play a role in the variation of prevalence rates.

As reported by Hinton and Lewis-Fernández (2011), studies demonstrating cross-cultural variations of posttraumatic symptomatology were considered in the development of the new DSM-5, and the recommended changes to the PTSD diagnosis were based on the results of these studies. As they noted, the decision to have a threshold of only one symptom for the distinct avoidance cluster was, in part, based on the lower rate of avoidance reported in certain cultures. The additions of a new section on culture-related diagnostic issues illustrate other efforts made to ensure that the DSM-5 classification of PTSD is both culturally sensitive and culturally appropriate (Lewis-Fernández, Hinton, & Marques, 2014).

In addition to studies examining the PTSD construct across different cultures, there have been studies that define posttraumatic responses according to more emic approaches. For instance, some researchers have found distinctive posttraumatic symptom expressions in different cultures that are quite dissimilar to the symptoms of PTSD. One of these cultural trauma syndromes is *baksbat* or "broken courage" in Cambodia (Chhim, 2012), which can include symptoms such as an inability to trust others, becoming more submissive, feelings of cowardliness, and becoming deaf and mute. Another example of a cultural syndrome is *ihahamuka* among Rwandan genocide survivors which is mainly characterized by a shortness of breath (Hagengimana & Hinton, 2009).

In summary, researchers are still trying to understand whether, when, or how posttraumatic psychological symptoms are more indicative of a culturally bound disorder, a universal construct, or something in between. Although further studies are warranted to clarify this situation, it is important to remember that these symptoms affect thousands every year and

that there is evidence to support the cross-cultural applicability of the PTSD construct (e.g., de Jong et al., 2001). Accordingly, the development of both appropriate instruments to assess PTSD and evidence-based treatments to treat the condition should be a high priority.

The assessment of PTSD. Accurate assessment of PTSD can be challenging, as researchers and clinicians rely mainly on subjective reports of symptoms, and the overlap of PTSD symptoms with other psychopathologies can be difficult to tease apart (e.g., Brewin, Andews, & Rose, 2003). Similar to other disorders that have high rates of comorbidity, using multiple methods to assess trauma symptoms is highly recommended (Keane et al., 1985). These methods can include not only a thorough clinical history from the affected individual and/or collateral informants, but also structured diagnostic interviews, self-report measures, and psychophysiological measures.

Structured diagnostic interviews are often recommended for the assessment of PTSD due to higher diagnostic accuracy than self-report measures (Keane et al., 1985), and to their conformity with EBA standards. Although they are commonly used in research settings, they appear to be less commonly used in clinical settings (except for forensic practices) (Aboraya, 2009; Keane, 1995; Keane, Buckley, & Miller, 2003). Structured diagnostic interviews or clinician-administered instruments can be sections of a more comprehensive assessment instrument, such as the PTSD module of the Structured Clinical Interview for DSM-IV (SCID-IV; First, Spitzer, Williams, & Gibbon, 1996), or stand-alone tools such as the Clinician-Administered PTSD Scale (CAPS; Blake et al., 1990). A 2005 survey of trauma specialists found that these tools were the two most commonly used clinician administered instruments in clinical settings and were also among the most frequently used for research purposes (Elhai, Gray, Kashdan, & Franklin, 2005). In addition, in their review of instruments used to assess PTSD

using EBA guidelines, Keane and colleagues (2008) rated the CAPS and the SCID-IV as the most strongly recommended clinician-administered instruments for diagnosis. The CAPS was especially recommended after receiving excellent ratings across all psychometric categories (e.g., norms, internal consistency, content validity, clinical utility), and was also suggested for use for purposes of treatment monitoring and treatment outcome evaluation (Keane et al., 2008). Both of these structured interviews have been updated to reflect the changes in the DSM-5 (First, Williams, Karg, & Spitzer, 2016; Weathers et al., 2013).

Self-report measures can provide information about the trauma experienced by the individual and about the presence, frequency, and duration of PTSD and other trauma-related symptoms. Self-report instruments continue to be widely used, as they tend to take less time to complete and are less costly than structured interview schedules. Based on their survey results, Elhai et al. (2005) found that the most commonly used self-report measures in clinical settings were the Trauma Symptom Inventory (TSI; Briere, 1995), the PTSD Checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993), and the Posttraumatic Diagnostic Scale (PDS; Foa, Cashman, Jaycox, & Perry, 1997). Two of these self-report measures (i.e., PCL and PDS) were also Keane and colleagues' (2008) most highly recommended self-report instruments for diagnosis, treatment monitoring, and treatment outcome, based on Hunsley and Mash's (2008) EBA guidelines.

Another category of PTSD assessments includes psychophysiological measures which do not involve subjective reports but, rather, examine psychophysiological reactivity to trauma-relevant stimuli by measuring different physiological reactions, such as heart rate, skin conductance, or even cerebral activity through electroencephalography. Although there is evidence of their clinical utility (Keane et al., 1998), they are not easily accessible for most

clinicians as they are often expensive and require special training for their use (Orr, Metzger, Miller, & Kaloupek, 2004). Psychophysiological measures have also been criticized for their diagnostic accuracy, as studies have shown than approximately 40% of individuals who are diagnosed with PTSD do not manifest any physiologic reactivity (Orr et al., 2004).

Although using a multi-method approach to the assessment of PTSD is recommended, it is not always feasible. For instance, there may be time or financial constraints, especially in research contexts such as epidemiological surveys where large numbers of participants need to be assessed. Whether using multiple methods or not, a critical component in choosing an instrument is examining its psychometric properties for the group that is to be assessed. As discussed in the next sections, variations in the expression of posttraumatic symptoms across individuals with differing cultural backgrounds and the cultural utility of an instrument also need to be considered in selecting and using an instrument.

Assessment of PTSD across cultures. Both the diagnostic criteria of PTSD and psychological measures of PTSD have mainly been developed in Western and industrialized countries (Keane et al., 2008). Assuming that this Western-based symptomatology is universal, that is using a pseudoetic approach to trauma (Renner, Salem, & Ottomeyer, 2007), may fail to take into account the culture-specific idioms of distress and, therefore, might not be considered a truly evidence-based approach to assessment. In other words, a pseudoetic approach to the assessment of PTSD may increase the risk of measurement bias.

When it comes to using evidence-based trauma assessment instruments among different cultures, there are three options available (van de Vijver & Tanzer, 2004): (a) conducting a literal translation into a target language; (b) culturally adapting an existing instrument; or (c) developing a new one that is tailored for a specific culture. Although some researchers have

chosen this last option (e.g., Miller et al., 2006), developing an entirely new measure can be very costly and time-consuming, and would limit cross-cultural comparisons. It is also not feasible to develop standardized measures for every existing culture and sub-culture.

Most PTSD instruments found in the literature that have been used in diverse cultural groups were literal translations. For instance, the CAPS has been translated into several languages such as Bosnian (Charney & Keane, 2007), Swedish (Paunović & Öst, 2005), and German (Schnyder & Moergeli, 2002). With regards to self-report instruments, the PCL has been translated into various languages including, but not limited to, Spanish (Miles, Marshall, & Schell, 2008), Chinese (Wang, Su, Bi, Wei, & Mo, 2012), and Malay (Bahari, Alwi, Ahmad, & Saiboon, 2015). Although these translated instruments have evidence of rather good psychometric properties for their specific sample (e.g. Bahari et al., 2015), they may not fully capture all the symptoms related to posttraumatic stress within the target culture by not including culturally relevant idioms of distress.

The only PTSD instrument that was specifically designed to be culturally adapted and used across diverse cultural samples is the Harvard Trauma Questionnaire (Mollica et al., 1992). This option, that is incorporating an emic approach by including culturally relevant constructs to an existing instrument, appears to be more efficient and has long been proposed as yielding the most accurate assessment possible (e.g., Berry, 1989). However, the overall psychometric properties of the HTQ have not been evaluated to support this claim. Indeed, adapting cross-cultural instruments for the assessment of posttraumatic stress symptoms comes with many challenges and many factors need to be considered to optimize their measurement accuracy. For adapted instruments that have already been used in research, a meta-analytic examination of the psychometric properties of these instruments can provide essential information regarding their

characteristics and likely value when used in subsequent research. In addition, because the assessment of traumatized individuals can have major clinical and societal implications (e.g., access to treatment, obtaining disability compensation, obtaining refugee status), an in-depth analysis of instruments designed for use in these evaluations is of the utmost importance.

The Harvard Trauma Questionnaire

The Harvard Trauma Questionnaire (HTQ) was originally developed by the Harvard Program in Refugee Trauma (HPRT) and the Indochinese Psychiatry Clinic in Massachusetts after years of extensive research and clinical experience with refugee populations (Mollica, McDonald, Massagli, & Silove, 2004). After unsuccessful attempts at assessing trauma with an Indochinese population using another standard instrument, staff at the clinic decided to create the HTQ by following the format of the Indochinese versions of the Hopkins Symptom Checklist-25 (HSCL-25; Mollica, Wyshak, de Marneff, & Lavelle, 1987), an instrument that assesses symptoms of anxiety and depression. The HTQ was developed as a cross-cultural, clinician-administered instrument to assess trauma and torture related to mass violence and their psychological impacts. It was intended to be used with clinical and community refugee populations, in both research and clinical settings (Mollica et al., 2004). Although the developers initially recommended its use for refugee populations, they have also used the HTQ among non-refugees (e.g., Silove et al., 2007).

The HTQ was the first cross-cultural trauma assessment instrument to be created and validated across various cultural groups. The first versions of the HTQ were validated among three Indochinese refugee populations: Cambodian, Vietnamese, and Lao (Mollica et al., 1992). Mollica and colleagues (1994) specified that the HTQ should not simply be translated into another language and then administered to traumatized individuals. Instead, they recommended a

rigorous adaptation and revision process supported by detailed knowledge of the culture, relevant life events of those who would complete the measure, and culture-specific symptoms for each new form of the HTQ and for each distinct group of trauma and torture survivors. In the HTQ manual, Mollica and collegues (2004) described evidence-based approaches to the cross-cultural adaptation of instruments that can help researchers and clinicians develop their own versions of the questionnaire. These adaptation procedures included essential components of cross-cultural adaptation such as guidance on establishing a group of experts to consult on the adaptation process, gathering qualitative information on cultural expressions of distress, translating items, and evaluating cross-cultural equivalency. They particularly noted the importance of cross-cultural equivalency by describing Flaherty et al.'s (1988) five equivalence dimensions (i.e., content, semantic, technical, criterion, and conceptual equivalence), which are similar to the ones described in the previous section. They also emphasized the use of focus groups to gain a better understanding of the cultural meaning and symptomatology of traumatic experiences of a specific group.

Systematic reviews of the literature have shown that the HTQ is commonly used in studies among refugee populations (Gagnon et al., 2004; Hollifield et al., 2002), and can be useful to exclude PTSD non-cases in specific forcibly displaced populations (Nakeyar & Frewen, 2016). The HTQ has also been recommended as a tool to be used across cultures by trauma experts (e.g. Keane et al., 2008). However these assertions were not supported by sufficient empirical evidence that would justify qualifying it as a culturally sensitive measure of PTSD symptoms. Nor have evidence-based assessment guidelines such as Hunsley and Mash's ever been used to evaluate the HTQ and support these claims. Because of these factors, it is imperative that a thorough review of its psychometric properties be conducted. The results of a

review could provide both evidence for the scientific status of the HTQ and also an indication of what is possible for the cultural adaptation of PTSD measures in general.

In the following paragraphs I will describe the structure of the HTQ and the range of populations with which it has been used. Although the different sections of the instrument vary among the various versions of the HTQ, four general sections can be found in most versions. Because of the need for the instrument to be culturally appropriate, the number and content of items of each section can also vary across versions. Part I includes items that describe a range of traumatic experiences, such as "combat situation" and "forced separation from family members." There are four possible responses for each item: (a) experienced, (b) witnessed, (c) heard about, and (d) no. The respondent is asked to indicate which of these responses (one or more) best represent his or her level of exposure to the trauma. Part II includes two open-ended questions related to the subjective experience of the respondent's most traumatizing event. Part III inquires about direct head injury and other events that could possibly lead to brain damage (e.g., starvation, suffocation). Part IV assesses posttraumatic symptoms, and it is this part that is the focus of the dissertation studies.

In Part IV, the first 16 items are derived from the DSM-III-R/IV PTSD criteria (APA, 1987, 1994) and are the same in every version of the HTQ. I refer to these 16 items as the HTQ-16 throughout the dissertation. The items represent the criteria for the intrusion/re-experiencing, avoidance/numbing, and hypervigilance/arousal symptom clusters. Sample items include: "recurrent thoughts/memories," "hard to concentrate," and "can't feel emotions." The respondent is asked to answer each item according to the following scale: 1 = "not at all," 2 = "a little," 3 = "quite a bit," 4 = "extremely." They are also asked to evaluate how much each symptom bothered them within the past week. The next set of PTSD symptom items are culture-

specific questions and, therefore, are tailored for each different version. For instance, the Arabic version that was validated among Iraqi refugees in the USA (Shoeb, Weinstein, & Mollica, 2007) consists of a total of 45 trauma symptom items and includes unique cultural items such as "Nafseetak ta'bana" or tired soul. As a further example, the Japanese version developed for earthquake survivors includes the cultural item "Disappointment upon awakening that life is not better," which is not found in other versions. The English version of the questionnaire includes 14 cultural symptoms that are based on clinical observations among Indochinese refugees resettled in the USA (i.e., literal translation of the Indochinese versions). This version has been commonly used in the empirical literature.

Scoring the trauma symptoms part of the HTQ includes calculating the mean item score for both the HTQ-16 and the full symptom scale (i.e. 16 DSM symptom items and the cultural symptom items). In both cases the mean item score is obtained by summing the individual item scores and then dividing by the number of items, yielding a mean item score for each scale. The range of possible scores for each scale is 1 to 4. To establish clinical caseness, Mollica and colleagues (2004) proposed two methods: (a) cut-off scores and (b) an algorithm approach. They provided instructions as to how to calculate a cut-off score for a specific group of respondents for three purposes: (a) screening, (b) clinical care, and (c) research. Although the community sample cut-off scores of 2 (Silove et al., 2007) and clinical cut-off score of 2.5 (Mollica et al., 1992) were established in some studies, the developers advised users to establish caseness benchmarks for each new population. In situations where the usual cut-offs are not applicable, and the establishment of new benchmarks is not possible, an algorithm approach is preferred. This approach is based on the DSM-III-R/IV criteria and it conforms to the well-established DSM guidelines for caseness. Specifically, to meet criteria for the diagnosis of PTSD, an individual

needs to score a minimum of 3 on one or more items of the intrusion/re-experiencing subscale, on three or more items on the avoidance/numbing subscale, and on two or more items of the hypervigilance/arousal subscale.

The developers of the original English and Indochinese versions of the HTQ have developed and validated other versions including: revised Cambodian (HPRT, 2000), revised English (HPRT, 1998), Bosnian (HPRT, 1998), Croatian (HPRT, 1998), Japanese (HPRT, 1996), Peruvian (HPRT, 2000), Kurdish (HPRT, n.d.), Arabic (HPRT, 2006), Spanish (HPRT, 2000), and Russian (HPRT, n.d.). The HTQ has also been translated and adapted for use in many other languages and for different cultural groups by other researchers (e.g., de Fouchier et al., 2012; Kleijn, Hovens, & Rodenburg, 2001), has been commonly used in research with refugees (Gagnon et al., 2004; Hollifield et al., 2002), and has been shown to be useful to exclude PTSD non-cases in certain forcibly displaced populations (Nakeyar & Frewen, 2016).

Although the measure is considered a gold standard for the cross-cultural assessment of posttraumatic stress symptoms, the psychometric properties of the HTQ symptom scales have only been examined in detail in a few studies beyond what has been reported in the instrument manuals. For example, Mollica and colleagues (1992) conducted a study on a sample of 91 Southeast Asian refugee patients (from Cambodia, Laos, and Vietnam) attending a psychiatric outpatient clinic. Good reliability properties were demonstrated for the 30 item symptoms scale (referred as the HTQ-30): interrater reliability for the trauma-related symptoms (k = .98); scale test-retest reliability with a 1-week interval (r = .89); and internal consistency (Cronbach α = .90). The sensitivity was found to be 78% and the specificity 65%, which indicate that the measure was better at identifying individuals within these cultural groups who have PTSD than at correctly identifying those who do not. Although this study showed evidence that the HTQ has

reasonably good psychometric properties for this population, the sample size was relatively small and the authors did not report the psychometric properties for the three cultural groups separately. Other investigators (Kleijn et al., 2001) examined the evidence of the psychometric properties of translations of the HTQ-16 in five other languages: Arabic, Farsi, Serbo-Croatian, Russian, and English. They reported that the internal consistencies of the different versions ranged from .74 to .89. This study also found that some of the items on the PTSD scale had low item-total correlations, which could potentially compromise the score reliability of the measure when used in other samples.

Some authors have suggested that the validity properties of the HTQ symptom scales are limited and that it may not be generalizable due to having been developed from a psychiatric outpatient population (Hollifield et al., 2002). In addition, it appears that the Cambodian version of the HTQ-16 may not be efficient at identifying true cases of PTSD among a non-refugee Cambodian community sample (Silove et al., 2007), although Mollica and colleagues (1992) reported more encouraging findings on this point. In a more recent study, Rasmussen and colleagues (2015) evaluated aspects of the construct validity of the HTQ-16 by examining its measurement invariance among 878 survivors of torture or other human rights violations who belonged to various cultural groups. Their results showed that the basic content validity, or overall structure of the PTSD construct as measured by the HTQ, was acceptable. However, there were significant response-style differences across cultures, as well as variations in the clinical thresholds of PTSD. Another study on the factor structure of PTSD using the HTQ-16 compared three models across three groups living in non-Western low- and middle-income countries (i.e. torture survivors in Northern Iraq, sexual violence survivors in the Democratic Republic of Congo [DRC], and Burmese refugees in Thailand) (Michalopoulos et al., 2015). The

factorial models examined were: (a) 3-factor DSM-IV model, (b) 4-factor "numbing" model, and (c) 4-factor "dysphoria" model. The results showed that all models had an adequate fit for the DRC and Burma samples. The 4-factor "numbing" model was the best fit for the DRC sample and the 4-factor "dysphoria" model for the Burma sample. None of the models had an adequate fit for the Northern Iraq sample. These important findings speak to the potential limitations of construct validity of the HTQ-16. Further research is also needed to evaluate other dimensions of its construct validity, such as convergent and discriminant validity.

Potential Moderators of Reliability and Validity

As mentioned previously, the data provided by an assessment instrument can be more reliable and valid for some populations than others. To obtain a clearer picture of the heterogeneous quality of this variance, it is important to examine its potential causes. In the event that the reliability and validity coefficients produced by the current RG and VG analyses show signs of heterogeneity (i.e. significant *Q* statistic), moderator analyses were conducted to assess the potential impact of sample and methodological characteristics on scores produced by the HTQ symptom scales, thus further examining the potential sources of variance. Commonly reported variables that have shown mean differences between groups with and without PTSD, or that have often been found to moderate reliability or validity estimates, were examined.

The sample characteristics examined as moderators included: (a) culture/ethnicity, (b) cultural context of the country of origin (individualist vs. collectivist), (c) type of trauma experienced, (d) being forcibly displaced, (e) age, and (f) gender. These moderator analyses were exploratory in nature due to the limited information available regarding their impact on the reliability and validity of HTQ scores. As described in detail throughout the introduction of this dissertation, there is evidence of the variation of PTSD symptomatology across cultures and

ethnicities (e.g., Kessler et al., 1999; Pole et al., 2008). Because of these findings and the fact that the studies using the HTQ were conducted across many cultures and ethnicities, this factor is considered a potential source of variance in the reliability and validity of HTQ scores.

As an extension of this potential moderator, the cultural background of the sample in terms of the construct of collectivism versus individualism was examined for its possible moderating effects. Individuals in collectivist cultures tend to be interdependent within their reference group (e.g. family, community, tribe), give more importance to the goals of this group, and behave according to the norms established within the group (Mills & Clark, 1982). On the other hand, people from individualist cultures are mostly independent from their reference group, give high priority to their own personal goals and behave according to their personal attitudes as opposed to the norms of their reference group (Triandis, 2001). Some studies have shown that a cultural group's orientation (i.e. collectivist or individualist) can have an impact on individual response styles on Likert-type questionnaires (e.g. Harzing, 2006), which may impact symptom endorsement. Specifically regarding posttraumatic stress, research has shown that individuals that come from more independent cultures and who have developed PTSD are more likely to have a change in self-definition and personal identity following a trauma that individuals from more interdependent cultures (Jobson & O'Kearney, 2008). In addition, Elsass (2001) found that, among South American cultures, individualist cultures tended to view PTSD avoidance symptoms as maladaptive coping, whereas collectivist cultures tended to consider avoidance symptoms to be adaptive. Such differences in perspective may have an impact on symptom reporting, and consequently, on score reliability. Because this qualitative study included a small sample, it only offers preliminary findings. Larger quantitative studies are needed to further examine the role of collectivism and individualism on the psychological effects of trauma. To

categorize cultures as collectivist or individualist in the current study, Hofstede's dimensions of national culture was used (Hofstede, Hofstede, & Minkov, 2010). As a result, only the countries in Hofstede's research studies were categorized and included in the moderator analyses using this cultural distinction.

Another variable that may contribute to the variance of the psychometric evidence of the HTQ is the type of trauma experienced by research participants. This variable was considered because the prevalence rate of PTSD varies according to the type of trauma experienced by individuals. For instance, the National Comorbidity Survey revealed that rape, childhood physical abuse, and childhood neglect were the traumas the most related to PTSD (Kessler et al., 1999). In a more recent study, the probability of developing PTSD after experiencing a sexual assault was 40.2%, as compared to 9.6% for severe accident, 9% for unexpected death of someone close, and 5.1% for natural disaster (Breslau, Troost, Bohnert, & Luo, 2013). The types of traumas in the RG and VG data were merged into two broader categories (i.e., intentional trauma and non-intentional trauma). Studies have shown that individuals who are exposed to intentional traumas had worse health outcomes than did individuals who experienced traumas that were not deliberate, such as natural disasters or accidents (Lange et al., 2003; Matthieu & Ivanoff, 2006; Van der Velden et al., 2006).

Being a forcibly displaced person (FDP), which includes refugees, asylum seekers and the internally displaced was also considered as a potential moderator. These individuals flee their home to escape traumatic events such as war, civil conflict, or persecution. The impact of these pre-migration stressors can also be compounded by resettlement difficulties or other post-migration problems such as detention (Ichikawa, Nakahara, & Wakai, 2006; Porter & Haslam, 2005). Researchers tend to find higher rates of psychological difficulties among individuals who

have survived war and have been displaced as compared to non-refugee groups (Fazel et al., 2005; Porter & Haslam, 2005; Steel et al., 2009). However, research has also found that estimates of PTSD prevalence among refugee populations can vary greatly, from 4% (Hauff & Vaglum, 1994) to 86% (Carlson & Rosser-Hogan, 1991).

Age was considered a relevant potential source of variance as some studies have found that rates of PTSD can differ according to age. For instance, Kessler and colleagues (1999) found in the National Comorbidity Survey in the USA that there were age differences in the prevalence of PTSD according to gender. Among women, there was no significant association between age and PTSD. Among men, there was a significant association between age and PTSD, however, there was no association after controlling for trauma exposure or risk of PTSD after trauma among the different age cohorts. Furthermore, analyses conducted on data from this survey indicated that the conditional risk for PTSD declined as age increased (Kessler et al., 1995; Norris, 1992).

A final sample characteristic that was considered as a potential moderator is gender, as it has also been frequently studied as a risk factor for the development of PTSD and may play a role in any heterogeneity of variance. Overall, the main findings of epidemiological research show that, although women tend to experience fewer traumatic events, they are more likely to develop PTSD than are men (Breslau et al., 1998; Perkonigg, Kessler, Storz, & Wittchen, 2000). More specifically, women had a significantly higher prevalence of PTSD than did men in large national community surveys in the US (Kessler et al., 1999; Pietrzak et al., 2011). These findings have also been replicated in other parts of the world (e.g., Creamer & Parslow, 2008; Ditlevsen & Elklit, 2010).

The methodological variables included as potential moderators were: (a) type of sampling

used (general community sample; sample seeking mental health services; sample seeking medical treatment), (b) the study sample was one for which the HTQ was validated, and (c) type of adaptation procedure used. The moderator analyses of the type of sampling variable were exploratory, and the analyses of the other two variables were hypotheses driven.

The sites from which the samples were recruited were examined. Psychometric properties of instruments can be dependent on the severity of the disorder measured (Haynes et al., 2011) and, as symptoms are generally more severe in individuals who are seeking mental health services or medical treatment than they are in individuals from non-clinical samples, the effect of recruitment site (i.e., clinical, community, or mixed clinical and community) were considered.

In previous sections of this introduction I discussed the important notions of equivalency when adapting instruments for different populations or purposes. Although Mollica and colleagues (2004) also explicitly highlighted this notion and provided recommendations with regards to adaptation procedures for other researchers in their manual, early in the article search, I noticed that many authors used original versions of the HTQ with populations that had not undergone a previous validation process. To illustrate, Gorst-Unsworth and Goldenberg (1998) used the HTQ symptom scale with a sample of male Iraqi refugees resettled in the United Kingdom. When they described their use of the HTQ, they stated only that it was administered in Arabic and Kurdish with the help of an interpreter. There was thus no indication that the researchers followed any adaptation guidelines. Internal consistency and convergent validity estimates should be highest in studies involving samples for which the HTQ was originally intended to be used, and discriminant validity estimates lower in such samples. Accordingly, moderator analyses were conducted to determine whether having a study sample that was culturally similar to the samples used in the development of the HTQ affected reliability and

validity estimates.

The Current Studies

The assessment of PTSD is a rather challenging endeavour due to the continuously evolving conceptualization of the construct and the absence of clear differentiation from other disorders. The cross-cultural variations in posttraumatic symptom expression further add to this complexity, thus making the assessment of PTSD across diverse cultural groups an even more challenging task. The various sources of cultural bias and measures to address them further complicate the development and validation of instruments, making the evaluation of the psychometric properties of these instruments highly imperative.

The available evidence regarding the psychometric properties of the HTQ symptom scales is inconsistent, and its overall psychometric characteristics across populations remain unknown. Because this instrument is intended to be used across cultures, a comprehensive analysis of the reliability and validity properties of this instrument across different populations is warranted. The current studies are the first attempt at aggregating empirical findings across studies and cultures to assess the cross-cultural applicability of the HTQ in light of evidence-based assessment guidelines. The results from these RG and VG meta-analyses will provide pertinent information that will help clinicians and researchers decide whether or not it is an appropriate assessment tool for their clients/participants of different cultural backgrounds.

Study 1

The Harvard Trauma Questionnaire: A Reliability Generalization Study

Abstract

We conducted reliability generalization (RG) meta-analyses of The Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992), a widely used cross-cultural instrument used to assess trauma and its psychological sequelae. A search of all publications and dissertations that used the symptom scales of the HTQ revealed that of the 384 studies considered for inclusion, only 44% of published studies and dissertations reported internal consistency estimates of their sample. In addition, only 16% of samples that would have required a cultural verification or adaptation of the scales were indeed administered a culturally adapted version. The random-effects RG included 183 reliability coefficients from 95 studies, representing data from 33,376 individuals. We characterized the variance of score reliabilities by considering sample and methodological variables that are predictive of the reliability coefficients. The mean reliability coefficients ranged from $\alpha = 0.74$, $p < 0.001$, 95% CI [0.72, 0.77] for the re-experiencing subscale to $\alpha = 0.93$, $p < 0.001$, 95% CI [0.92, 0.94] for the HTQ-30. Overall, the HTQ-16 and HTQ-30 are likely to provide reliable scores across a large diversity of populations. However, the re-experiencing, avoidance and arousal subscales should be used with caution due to lower overall mean reliability estimates and inadequate performance in certain samples. We discuss the importance of following journal article reporting standards and carefully implementing adaptation strategies in cross-cultural assessment.

Keywords: posttraumatic stress disorder, assessment, cross-cultural assessment, reliability, meta-analysis

The Harvard Trauma Questionnaire: A Reliability Generalization Study

Although traumatic events and posttraumatic stress disorder (PTSD) are considered prevalent in all cultures (Keane, Silberbogen, & Weierich, 2008), epidemiological research has found varying rates of the overall lifetime prevalence of PTSD across countries. Studies in the United States have shown that PTSD is present in 6.4% to 9.5% of the general population (Breslau, Davis, Andreski, & Peterson, 1991; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995; Pietrzak, Goldstein, Southwick & Grant, 2011). Based on data from the World Mental Health Surveys in 24 countries, the World Health Organization reported estimates for lifetime prevalence for PTSD ranging from 0.3% in China to 8.8% in Northern Ireland (Koenen et al., 2017). In addition, a systematic review on the prevalence of mental disorders in refugees who were resettled in Western countries showed that 9% of them suffered from PTSD (Fazel, Wheeler, & Danesh, 2005). Even though there are differences in rates of PTSD across studies, there is no question that the negative psychological effects of traumatic events are considerable.

Although the diagnostic criteria have changed across editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases (ICD), PTSD is commonly associated with symptoms such as re-experiencing, avoidance, numbing and hyperarousal. Despite the widespread occurrence of PTSD, there has been some criticism surrounding the cross-cultural applicability of the diagnosis. Some investigators have regarded some diagnostic criteria as reflecting Western cultural constructions and not being generalizable across cultures (e.g., Jones et al., 2003). For example, avoidance/numbing symptoms may be less pronounced in some cultures (Dyregrov, Gupta, Gjestad, & Mukanoheli, 2000; Marsella, Friedman, Gerrity, & Scurfield, 1996; McCall & Resick, 2003).

As psychodiagnostic assessments rely heavily on the conceptualization of the construct they are

intended to measure, it is not surprising that there has also been much disagreement about the cross-cultural applicability of screening tools used to assess PTSD. Psychological measures of PTSD have mainly been developed in Western and industrialized countries (Keane et al., 2008). Assuming that the symptomatology is universal (i.e., using an etic approach to trauma assessment) fails to take into account the culture-specific idioms of distress. Clearly, it is not feasible to develop standardized measures for every existing culture and sub-culture, but including culturally relevant constructs in the assessment is ideal for more accurate measurement (i.e., incorporating an emic approach) (e.g. Renner, Salem, & Ottomeyer, 2007). Researchers who choose to use an instrument with different cultural groups must be cognizant of the increased risk of various types of measurement bias, such as construct, method (sample, instrument, response styles, administration) and item bias (He & van de Vijver, 2012) and adapt the instrument accordingly. Establishing the various types of equivalence such as content, linguistic, semantic, conceptual, scale, technical, normative, and cross-cultural equivalencies can help minimize bias in the cross-cultural adaptation process (Keane, Kaloupek, & Weathers, 1996). Because of these complexities in cross-cultural measurement that can impact accurate assessment, an examination of the psychometric properties of instruments that are intended to be used across diverse samples is of utmost importance.

The Harvard Trauma Questionnaire

The Harvard Trauma Questionnaire (HTQ) was developed by Mollica and colleagues (1992) in response to the need for culturally sensitive trauma measures. The HTQ is a clinician-administered questionnaire that is intended to be adapted and used across cultures, and has been validated by its developers for various cultural and linguistic groups (Shoeb, Weinstein, & Mollica, 2007). Most versions of the HTQ consist of four parts. Part 1 measures different

traumatic events, part 2 is an open-ended description of the most traumatic event experienced by the respondent, part 3 evaluates the circumstances surrounding possible head injury and, finally, part 4 is a list of trauma symptoms. The first 16 items (HTQ-16) of this last part are derived from the DSM-III-R/IV (American Psychiatric Association [APA], 1987, 1994) PTSD criteria, and they are the same in every version of the HTQ. A second set of trauma symptom items in part 4 are culture-specific questions and are tailored for each version of the HTQ. The HTQ has also been translated and adapted in many other languages by other researchers (e.g., Kleijn, Hovens, & Rodenburg, 2001). Systematic reviews of tools used to assess the health of refugees have found that the HTQ has been extensively used by researchers to assess trauma and its sequelae (Gagnon, Tuck, & Barkun, 2004; Hollifield et al., 2002), and can be useful to exclude PTSD non-cases in certain forcibly displaced populations (Nakeyar & Frewen, 2016). In addition, in their review of PTSD measures, Keane et al. (2008) recommended the use of the HTQ for the assessment of PTSD across cultures. They described the HTQ as having "linguistic equivalence across the many cultures and languages with which it has been used thus far" (p. 297). However, they did not provide sufficient empirical evidence to support the cross-cultural applicability of the instrument or evaluate it according to established evidence-based assessment guidelines, such as Hunsley and Mash's (2008).

Although the measure is frequently used and appears to have a reputation of being a good example in terms of cross-cultural trauma assessment, the psychometric properties of the HTQ have only been examined in a small number of studies. For example, Mollica and colleagues (1992) conducted a study on a sample of 91 Southeast Asian refugee patients (from Cambodia, Laos, and Vietnam) attending a psychiatric outpatient clinic. Evidence of good reliability of the HTQ-30 (16 DSM PTSD items and 14 Indochinese cultural items) was demonstrated in this

sample, including interrater reliability (k = .98), test-retest reliability (r = .92), and internal consistency (Cronbach α = .96). With respect to the DSM-III-R PTSD diagnosis based on semi-structured clinical interviews, sensitivity was found to be 78% and the specificity was 65% for the HTQ-16. When the 14 cultural symptom items were added, the sensitivity remained the same, whereas the specificity increased to 72%. The inclusion of the cultural symptoms thus improved the ability of the scale in accurately detecting PTSD. Although this study showed that the HTQ symptom scales had reasonable psychometric properties with this sample, the authors did not report the reliability coefficients for the three cultural groups separately and the total sample size was relatively small. Kleijn and colleagues (2001) examined the psychometric properties of translations of the HTQ in five languages: Arabic, Farsi, Serbo-Croatian, Russian, and English. They reported that the internal consistencies of the different versions of the HTQ-16 ranged from .74 to .89. This study also showed that some of the items on this scale had low item-total correlations, which could potentially compromise the score reliability of instrument when used with other samples. Further to this point, the measurement invariance of scores on the HTQ-16 across various populations has been found to be questionable (Rasmussen, Verkuilen, Ho, & Fan, 2015).

Although these studies showed that the scores obtained on the HTQ were, at minimum, adequately reliable according to Hunsley and Mash's (2008) benchmarks, its overall psychometric properties with different cultural groups remains unknown. Because this instrument is intended to be used across cultures, a further analysis of studies that have used the HTQ in different populations is warranted. Aggregating empirical findings across studies and cultures will provide pertinent information regarding the potential cross-cultural utility of the

HTQ, thereby helping clinicians and researchers decide whether it is an appropriate assessment tool for use with individuals of different cultural backgrounds.

Reliability and Reliability Generalization

Reliability, or the precision of a measure, represents "the consistency of the scores across instances of the testing procedures" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 33). In other words, reliability estimates represent the proportion of variance that is explained by the true score itself (i.e., characteristics measured in the test), as opposed to the part of the variance that is explained by sampling error or measurement error (Graham, Yenling, & Jeziorski, 2006). As the reliability of test scores also sets a limit on the possible validity estimates of an instrument, examining the degree of consistency of scores is imperative in determining if the instrument is indeed scientifically sound.

It is also important to note that reliability refers to the properties of test scores of a specific sample, and not to the properties of an instrument itself. Numerous factors, including sample characteristics and the context in which the assessment occurs, can influence the degree of consistency of test scores. For instance, the same instrument can produce scores with differing reliability when administered to more or less homogenous samples (Thompson, 1994). In sum, reliability is not only dependent on the quality of a measure, but also the characteristics of the sample, sample size, and administration and scoring procedures (Barnes, Harp, & Jung, 2002).

Researchers often assume that reliability estimates of prior studies are directly generalizable to data collected in their own studies (Vacha-Hasse, Kogan, & Thompson, 2000). The common error of referring to the reliability of an instrument as an inherent characteristic of the instrument may be responsible for researchers ignoring examining the score reliability of

their sample, which may then lead researchers to unwittingly interpret unreliable data (Graham, Diebels, & Barnow, 2011; Henson, Kogan, & Vacha-Haase, 2001). This tendency to generalize the reliability properties across studies has been called "reliability induction" by Vacha-Haase et al. (2000). These authors argued that reliability induction is only plausible if the previous samples have similar sample composition and score variability to the current set of participants. Even when samples may seem to be comparable, examining score reliability is still important as there may be subtle differences between the samples.

As reliability estimates vary across samples and different administrations of a given measure, Vacha-Haase (1998) employed a meta-analytic method called "reliability generalization" (RG) to examine the likely generalizability of reported score reliabilities of a specific measure across studies. When conducting a RG, all studies that have provided a reliability estimate of their sample are retrieved and coded. This collection of estimates is used to produce a mean reliability score and to examine the variability of score reliabilities. When the variability across estimates is significant, the RG can also be used to identify which sample and study characteristics account for the variations in score reliability for the measure (e.g., sample size, country of study, mean age of sample). RG results such as the mean reliability estimate, confidence intervals around the mean reliability estimate, and the identification of moderator variables that account for variations in score reliability provide useful information to clinicians and researchers when selecting an assessment tool appropriate for a specific task.

The Current Study

Considering the wide range of score reliabilities for the HTQ symptom scales reported in development and validation studies of the measure, the present research was designed to evaluate the extent to which the original and adapted versions of the HTQ symptom scales (and subscales)

produced reliable scores in the research literature. The internal consistencies of these scales, across samples, were investigated by using a reliability generalization analysis. This meta-analytic method determines the average reliability of scores obtained on the HTQ symptom scales and subscales, and also accounts for variability in score reliability across samples through moderator analyses.

Specifically, moderator analyses were conducted to assess the potential impact of sample and methodological characteristics on scores produced by the HTQ symptom scales and subscales, thus further examining the potential sources of variance. Commonly reported variables that have shown mean differences between groups with and without PTSD, or that have often been found to moderate reliability estimates, were examined. With regards to sample characteristics, age was considered a relevant potential source of variance as some studies suggest that the rates of developing PTSD after exposure to a traumatic event may decline as age increases (Kessler et al., 1995; Norris, 1992). Gender was also included as a moderator, as epidemiological research shows that, compared to men, women tend to experience fewer traumatic events but are more likely to develop PTSD (Breslau et al., 1998; Perkonigg, Kessler, Storz, & Wittchen, 2000).

Because a main purpose of this study was to evaluate the cross-cultural applicability of the HTQ, and considering that there is evidence of the variation of PTSD symptomatology across cultures (e.g., Kessler et al., 1999; Pole, Gone, & Kulkarni, 2008), the impact of the cultural background of study participants of the study was examined. As an extension of this potential moderator, the cultural background of the sample in terms of the construct of collectivism versus individualism was examined for its possible moderating effects. For example, Elsass (2001) found that, among South American cultures, individualist cultures tended to view PTSD

avoidance symptoms as maladaptive coping, whereas collectivist cultures tended to consider avoidance symptoms to be adaptive. Such differences in perspective may have an impact on symptom reporting, and consequently, on score reliability. To categorize cultures as collectivist or individualist in the current study, Hofstede's dimensions of national culture were used (Hofstede, Hofstede, & Minkov, 2010).

Other sample characteristics that were included as possible moderators were type of trauma and whether individuals had been forcibly displaced. The types of traumas were grouped into two categories: intentional trauma (e.g., war, torture, sexual assault) and unintentional trauma (e.g., natural disasters, accidents). These categories were chosen because studies have shown that individuals who are exposed to intentional traumas had worse health outcomes than did individuals who experienced traumas that were not deliberate (Lange et al., 2003; Matthieu & Ivanoff, 2006; Van der Velden et al., 2006). Whether study participants were forcibly displaced was also included as a moderator variable, as researchers tend to find higher rates of psychological difficulties among individuals who have been displaced as the result of war as compared to non-displaced groups (Fazel et al., 2005; Porter & Haslam, 2005; Steel et al., 2009).

Two methodological variables were included as potential moderators. Psychometric properties of instruments can be dependent on the severity of the disorder measured (Haynes, Smith, & Hunsley, 2011) and, as symptoms are generally more severe in individuals who are seeking psychotherapy or medical treatment than they are in individuals from nonclinical samples, the effect of recruitment site (i.e., clinical, community, or mixed clinical and community) was considered.

Finally, it was predicted that internal consistency estimates would be highest in studies involving samples for which the HTQ was originally intended to be used. Accordingly,

moderator analyses were conducted to determine whether having a study sample that was culturally similar to the samples used in the development of the HTQ affects reliability estimates.

Method

Literature Search and Data Collection

To identify studies for possible inclusion in the RG, a literature search of studies having used the HTQ was performed in the PsycINFO, PubMed, PILOTS (Published International Literature on Traumatic Stress), and Web of Science databases. The search strategy was based on a cited reference search, and enabled the identification of both published studies (i.e. article or book format) and unpublished dissertations. Most authors who use the HTQ cite Mollica and colleagues' first published article describing the development and initial psychometric property evaluation of the measure (Mollica et al., 1992). Some authors also cite the HTQ manual, which was published a few years later and includes a more detailed description of the measure (Mollica et al., 1996). Therefore two separate cited reference searches were conducted. The studies collected were published within a timeframe ranging from the year 1992 (the year of the first published study of the HTQ) to September 2016, when the final literature search was performed.

Inclusion criteria. Studies meeting the following inclusion criteria were selected: (a) the authors used at least one of the HTQ symptom scales; (b) the study was empirical offering quantitative data on the HTQ symptom scales (c) the sample consisted of participants 7 years of age or older (to be consistent with DSM-5 [APA, 2013] criteria); (d) the study was written in English or French (languages spoken by the authors); (e) the study was accessible through our university's library network (including inter-library loans); and (f) the study was available in article, book chapter, or dissertation format.

Data reduction. The cited reference search of Mollica et al. (1992) yielded a total of 1,505 entries. The cited reference search of Mollica et al. (1996) yielded 66 entries. There was a final total of 1,571 entries for all searches combined. More specifically, there were 126 entries in PsycINFO, 125 in Medline, 724 in PILOTS, and 596 in Web of Science. After removing duplicate entries, 862 unique studies were retrieved for further review. The full text of the selected studies was read by the first author and retained if the study fit the inclusion criteria. This detailed examination reduced the number of potential studies to 384. Of the excluded studies, 41% were not empirical quantitative studies, 35.8% did not use the HTQ symptom scales, 0.4% did not use a sample of participants aged 7 years and older, 4.6% were not written in English or French, 12.1% were not accessible within our university library network, and 6.1% were not available in article, book chapter, or dissertation format (see Figure 1).

Once all the studies considered for inclusion were gathered, the first author (a clinical psychology doctoral student with previous coding experience) used a coding manual to code the journal discipline and the reliability mentioning practices of the studies' authors. Journal discipline included four possible categories: (a) medical; (b) psychiatry; (c) general mental health; and (d) other. Although "psychology" was initially considered as a possible category, journals of this specific discipline were difficult to distinguish from other general mental health journals and were therefore included in this latter category. The manner in which authors commented on the reliability properties of the HTQ symptom scales included four possible options: (a) provided a reliability coefficient from their specific sample; (b) reported a reliability coefficient from another study; (c) mentioned that the measure is reliable without providing a value; and (d) made no reference to score reliability.

Authors who did not report Cronbach's alpha of their samples and who provided their email addresses were contacted and asked to either provide the coefficient alpha for their sample or to send their data sets to calculate the coefficient. Out of the 108 authors who were contacted, 35 responded (32.4% response rate), and 18 provided the internal consistency coefficient for their sample. Independent t-tests revealed no significant differences between the mean published and unpublished reliability estimates for both the HTQ-16, $t(11) = 0.49$, $p = .63$, and HTQ-30, $t(4) = -2.52$, $p = .65$.

Prior to conducting the RG analyses, multiple publications on the same data set were identified to ensure the statistical independence of the studies to be included in the analyses. Most multiple publications on the same data set were easily identifiable but, to ensure that all were identified, the detection heuristic developed by Wood (2008) was used as a guideline. This method included verifying if one or more of the "same" authors were included in the suspected multiple publications and determining whether the study and sample characteristics were similar in these studies. When deciding which study to eliminate from further use, several factors were considered, such as the amount of descriptive data reported and the reporting of the coefficient alpha of the sample. In addition, when studies provided alpha coefficients of different samples within the same study, each sample was recognized as separate and was retained for inclusion in the analyses.

In summary, 384 studies matched the inclusion criteria of which 185 studies had sample reliability estimates, either reported in the publication or provided by the authors. After removing studies using the same samples, studies that included modified versions of the HTQ symptom scales (i.e., added or subtracted items), and unusable reliability information (such as reporting only the range of alpha across subscales), there were a final total of 95 studies and 111 samples

to be included in the analyses. Specifically, this included 36 alpha coefficients for the HTQ-30 (i.e., the 16 DSM items and the 14 Indochinese culture-specific items), 70 for the HTQ-16, 22 re-experiencing subscale, 27 for the avoidance/numbing subscale, and 28 for the arousal subscale.

Coding of descriptive data and moderators. Once all the relevant articles were identified, internal consistency values, sample characteristics, and methodological characteristics were coded by the first author for the HTQ symptom scales and subscales. A coding manual (see Appendix A) was developed to help guide the coding process and data were entered on a coding sheet (see Appendix B). The coding manual and the selection of coding variables were based on a review of the literature on trauma, cross-cultural measurement and reliability generalization. The first author initially developed the manual and then revised the document in collaboration with the second author. Some variables initially considered (e.g., level of education, length of stay in host country/area) were omitted due to the lack of reporting of such data, or because the data were presented in different incomparable formats. The study and sample variables were used in the primary descriptive statistics of the overall pool of studies that used the scales and subscales. Some of these variables were also examined as potential moderators. Study and methodological characteristics included the following categorical variables: (a) language of administration of the HTQ; (b) country of study; (c) type of sample (community; seeking mental health services; seeking medical treatment sample); (d) version of the HTQ (original or adapted version), (e) whether an adaptation or cultural verification was needed for the sample (yes or no); (f) adaptation procedure used (cultural adaptation only, translation only, or both); (g) adaptation procedure followed the developers' recommendations (yes or no), (h) adaptation procedure followed other experts' recommendations (yes or no); (i) translation procedure used (oral translation; forward translation; back translation; blind back translation); (j) a consensus

approach was used during adaptation (yes or no); (k) the adaptation included a pre-test (yes or no); and (l) type of administration (clinician administered; non-clinician administered; self-report).

Sample characteristics included sample size, the mean age of participants, and gender composition of the sample (percentage of males), which were coded as continuous variables. Sample characteristics also included the following categorical variables: (a) country of origin of participants; (b) country of origin's cultural orientation (individualist or collectivist) based on Hofstede's categorization (only countries classified by Hofstede were coded); (c) type of trauma experienced (intentional [e.g., war, torture, sexual assault] or unintentional [e.g., natural disasters, accidents]); and (d) sample consisted of forcibly displaced persons (yes or no).

To assess the reliability of the coding procedures, the second author (a clinical psychologist and professor with extensive experience in meta-analysis) coded 20% of the studies that were included in the meta-analyses. Reliability analyses were performed using SPSS version 20. This step was important to help eliminate random variation in the analysis and increase the reliability and power of the results (Dieckmann, Malle, & Bodner, 2009). The inter-coder reliability was calculated by using the kappa (*k*) statistic for the categorical variables, which is a measure of the agreement between two raters (Cohen, 1960). Specifically, the *k* statistic calculates the proportion of agreement while controlling for the proportion that would occur by chance. The results had a rating of "good" according to Hunsley and Mash's (2008) criteria, with a mean $k = 0.76$. The inter-coder reliability of continuous variables was calculated by using intraclass coefficients (ICC), (Shrout & Fleiss, 1979) and were also considered "good" as assessed with Hunsley and Mash's benchmarks for ICC, with a coefficient of 0.85. Several discrepancies were caused by unclear reporting of data in the studies, vague descriptions of the

versions of the HTQ used, and vague descriptions of any adaptation procedures that were used. Discrepancies between the coders were discussed until a unanimous agreement was reached.

Data Analysis

Calculating mean effect sizes. The first step of the analyses was to enter the reliability information of each sample (sample identifier, sample size, and reliability coefficient) into the Comprehensive Meta-Analysis software version 2.2.064 (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005). As reliability estimates such as Cronbach's alpha are variance accounted-for statistics, they are often considered equivalent to r^2 (Thompson & Vacha-Haase, 2000). Prior to being entered into the CMA software, these coefficients were converted into their square root, the same metric as a correlation. To compensate for the high level of skewness and nonnormal distribution usually found with correlations, a Fisher's *r*-to-*z* transformation was then performed, and all subsequent analyses were performed with these transformations (see Borenstein, Hedges, Higgins, & Rothstein, 2009).

Each sample was assigned a relative weight, the inverse-variance, which is determined primarily by the sample size (i.e., larger sample sizes are generally assigned more weight), and a 95% confidence interval. An overall mean effect size was then calculated, including its statistical significance. A calculation of the standard error of the mean effect size then allowed the computation of the upper and lower 95% confidence intervals for the mean effect size. The Fisher's *z* values were then converted back to their original internal consistency values to facilitate interpretation.

This meta-analytic procedure was performed to produce mean internal consistency values for each of the following scales: (a) HTQ-30; (b) HTQ-16; (c) re-experiencing subscale; (d) avoidance/numbing subscale; and (e) arousal subscale.

Random-effects model. A random-effects model was chosen to account for the variations of the true effect from sample to sample. This model allows for the true effect to be different across samples, in contrast to the fixed-effect model that assumes that the samples have a common true effect and that the differences in observed effects are due entirely to sampling error alone (Borenstein, Hedges, Higgins, & Rothstein, 2010). Thus, the random-effects model considers that the effect sizes may come from different heterogeneous populations. Considering the vast differences in cultural origin, language spoken, trauma experienced, and other heterogeneous characteristics of the samples to include in the analyses, a random-effects model was considered the most suitable for this meta-analysis.

Heterogeneity of effects. Following the procedures of recent reliability generalization studies (e.g., Graham et al., 2011; Therrien & Hunsley, 2013), a *Q* test of homogeneity (Cochran's *Q*; Cochran, 1954) and I^2 index (Higgins & Thompson, 2002) were used to evaluate the degree of dispersion of reliability estimates around the mean reliability coefficient for each scale or subscale. A significant *Q* statistic indicates that there is true effect size variance among the studies and suggests that moderator analyses should be performed to ascertain the sources of heterogeneity. Although the *Q* statistic evaluates the null hypothesis that the studies included in the meta-analysis have a common effect size, it does not provide information regarding the extent to which heterogeneity plays a role in the overall reliability estimate. An I^2 index was therefore calculated to assess the proportion of the observed variance that is due to true heterogeneity rather than being due to chance alone (Higgins & Thompson, 2002). To illustrate, an I^2 of 0 would indicate that the variability in effect sizes is due only to sampling error and not to between-study heterogeneity; a higher I^2 index implies that there is a higher proportion of variability that is caused by between-study heterogeneity. To help gauge the different levels of

heterogeneity and interpret the I^2 index, the following benchmarks proposed by Higgins and Thompson (2002) were used: 25% = low, 50% = moderate, and 75% = high.

Spearman-Brown Prophecy formula. Internal consistency estimates vary according to test length. The higher number of test items, the higher the internal consistency. Because the different symptom scales and subscales of the HTQ have different number of items, the Spearman-Brown Prophecy formula (Brown, 1910; Spearman, 1910) was used to estimate the impact of adding a number of equivalent items to the scales/subscales. For example, 14 items would need to be added to the HTQ-16 to compare it to the total 30-item scale. To compare the PTSD subscales with each other, the subscales for re-experiencing and arousal would each be increased to a total of 7 items, which is the number of items in the avoidance subscale.

Possibility of publication bias. To take into account the possibility that authors do not always report reliability coefficients in their study, a Fail-Safe *N* for RG (Howell & Shields, 2008) was calculated for each meta-analysis. This formula examines the possible influence that unpublished effects may have on the mean reliability values obtained in the meta-analyses. Funnel plots were not used to assess publication bias due to the risk of visual misinterpretation (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006).

Analysis of moderators. Moderator analyses are an important step when conducting an RG to clarify the sources of heterogeneity (Rodriguez & Maeda, 2006), with many RG studies finding statistically significant relations between reliability coefficients and various types of moderators (see Vacha-Haase & Thompson, 2011, for a review). Therefore, to examine possible sources of the heterogeneity of variance of the reliability estimates, an analysis of moderators was performed on scales that had statistically significant heterogeneity.

Because the potential moderators of this RG included both continuous and categorical variables, two different strategies were used to assess their role in accounting for heterogeneity. First, to examine the relations between continuous variables and score reliability, mixed effects (method of moments) meta-regression analyses were performed. These analyses determine whether the effect sizes are moderated, or predicted, by these variables. Second, for categorical variables, a series of random effect subgroup analyses were used to examine the relations between the overall internal consistencies of each scale and study and sample characteristics. Because some of the variables had an insufficient number of studies per subgroup (i.e., 2 or more), they were merged together to create broader categories to increase the power of the results. Specifically, "culture" was separated according to larger geographic cultural groups (e.g., Asian, European, African), and "types of trauma" were categorized as either "intentional trauma" or "unintentional trauma".

Results

The HTQ Symptom Scales in the Empirical Literature

Overall, 384 empirical articles, book chapters, and dissertations included a version of the HTQ symptom scales in a study. After removing studies using data from the same samples, there were 216 independent studies and 242 independent samples (67,360 individuals) that were administered the scale. The samples came from over 59 countries, representing over 88 cultures. The studies were conducted in over 61 countries and the scale was administered in over 74 different languages. See Table 1 for an overview of the HTQ adaptation and administration characteristics.

Reporting Practices of Reliability Estimates

Of the 384 studies considered for inclusion, 34.9% made no reference to score reliability, 7% reported a reliability coefficient from another study, and 13.5% mentioned that the measure was reliable without providing a value. Finally, 171 (44.5%) studies reported a reliability coefficient based on data from their specific sample. Of these, 167 specifically provided the Cronbach's alpha values for the HTQ symptom scale of their sample. Reliability reporting practices across journal type were variable, as 19.5 % of studies published in medical journals, 35% of studies in psychiatry journals and 51.7% of studies in general mental health journals reported sample-specific reliability estimates. These mediocre reporting results are similar to the rates found in other studies investigating psychometric reporting practices in education and mental health-related fields (Barry, Chaney, Piazza-Gardner, & Chavarria, 2014; Hall, Ward, & Comer, 1988; Qualls & Moss, 1996). This suggests that the tendency to not report psychometric data of sample scores is longstanding and does not appear to be changing despite the practice guidelines that have been put in place over the years, such as the Journal Article Reporting Standards (JARS) published by the American Psychological Association (American Psychological Association Publications and Communications Board Working Group on JARS, 2008).

Overview of Studies Included in the Meta-Analyses and Descriptive Statistics of Potential Moderators

Because fewer than half of the studies provided reliability estimates, the present meta-analyses included 111 samples from 95 studies, representing data from 33,376 individuals. See Appendix C for the list of studies included in the RG analyses. The mean sample size was 301, the mean age of participants was 36 years, and, on average, 47% of study participants were male.

Table 2 presents information on these characteristics for each of the meta-analyses that were conducted. The studies were conducted in over 40 countries, with samples coming from over 42 different countries. Consistent with this, the HTQ symptom scales were administered in over 43 different languages. See Table 1 for an overview of the HTQ adaptation and administration characteristics.

Of the study samples that reported the sampling procedure (n = 108), 60.2% recruited participants from the community, 26.9% recruited participants who were seeking mental health services, 11.1% recruited participants who were seeking medical treatment, and 3.7% recruited from multiple sources. Forty-six percent of the samples were composed of forcibly displaced individuals (i.e., refugees, asylum seekers, internally displaced people) as opposed to natives or immigrants. The most common type of traumatic events was intentional traumas, with 78.2% of samples having all or most individuals who experienced these traumas. Unintentional traumas were experienced by 12.7% of the samples and the remaining 9.1% of samples consisted of individuals who had been exposed to both types of traumas.

Reliability Generalization Analyses

Descriptive statistics and results of the RG of each of the scales and subscales are presented in Table 3. The mean reliability coefficients ranged from $\alpha = 0.74$, $p < 0.001$, 95% CI [0.72, 0.77] for the re-experiencing subscale to $\alpha = 0.93$, $p < 0.001$, 95% CI [0.92, 0.94] for the HTQ-30. All scales and subscales had a significant Q statistic, indicating that the studies in each meta-analysis did not share a common effect size. In addition, the I^2 values ranged from 79.02% to 96.14%. The significant Q statistics and high I^2 indicate that there is an important amount of variance across samples for all RG, therefore further examination of potential study and sample characteristics that account for this variability is warranted.

Spearman-Brown formula. The full HTQ symptom scale has 30 items. If, to make it comparable in length to this scale, 14 items of similar content were added to the HTQ-16, the mean reliability coefficient would increase from .89 to .94. The avoidance subscale has 7 items: an addition of 3 items of similar content to the re-experiencing scale would increase the mean alpha from .74 to .83, and an addition of 2 items of similar content to the arousal subscale would increase the alpha from .79 to .84.

Publication bias. Based on a threshold internal consistency value of .7, the Fail-Safe *N* for obtained RG values for the scales and subscales ranged from 77 to 339. Although there are no specific guidelines regarding adequate Fail-Safe *N* values, in an example, Howell and Shields (2008) considered a Fail-Safe *N* of 60 to be robust to publication bias. Because all reliability estimates for the HTQ symptom scales and subscales included in the current RG are above 60, they are unlikely to be affected by publication bias.

Moderator analyses. As all five RG analyses had a significant *Q*-statistic, moderator analyses were conducted for each scale and subscale (see Table 4 for a summary of moderating effects). Samples that had missing information for a specific variable were excluded from these analyses. Age, being a sample for which the HTQ was validated, being forcibly displaced participants and the type of sampling were not significant for any of the five RG analyses. Gender was significant only for the re-experiencing subscale, $\beta = -0.003$, $Q_{model} = 4.66$, $p = 0.031$, 95% CI [-0.007, 0.000]. Specifically, study samples comprised of mainly female participants had higher reliability estimates. Subgroup analyses revealed that the summary effect sizes for "cultural group" were significantly different from each other for the re-experiencing subscale. "African cultures" had the highest mean Cronbach's alpha coefficient ($\alpha = 0.77$, $p < 0.001$, 95% CI [0.66, 0.85]) and "Asian cultures" the lowest ($\alpha = 0.56$, $p < 0.001$, 95% CI [0.40,

0.70]). “Cultural group” was not a moderator for the other scales. “Type of trauma” was also a moderator for the re-experiencing subscales. “Intentional traumas” had a significantly higher mean reliability coefficient ($\alpha = 0.78$, $p < 0.001$, 95% CI [0.75, 0.80]) than “unintentional traumas” ($\alpha = 0.72$, $p < 0.001$, 95% CI [0.67, 0.76]). “Type of trauma” was not significant for the other scales. Finally, the type of culture (individualist/collectivist) was a moderator for the avoidance subscale, with “Individualist country” ($\alpha = 0.78$, $p < 0.001$, 95% CI [0.75, 0.80]) being significantly higher than “Collectivist country” ($\alpha = 0.68$, $p < 0.001$, 95% CI [0.56, 0.77]). This variable did not have a moderating impact on the other scales.

Discussion

The main purposes of this study were twofold: (a) to gather updated descriptive information regarding the use of the HTQ symptom scales in the empirical literature, and (b) to evaluate the reliability properties of this measure by calculating mean Cronbach’s alphas and examining sample and methodological characteristics that may influence the internal consistency of each scale and subscale.

The HTQ in the Empirical Literature

In terms of the overall usage of the HTQ symptom scales (and subscales) in the literature, it is clear that they remain commonly used instruments among trauma researchers. In fact, since the most recent systematic review describing the popularity of the HTQ in 2004 (Gagnon et al., 2004), over 295 empirical studies have been published or written in dissertation format that included the HTQ symptom scales. The HTQ has also been used globally with a range of samples across a wide range of cultural and ethnic backgrounds, including refugee and non-refugees, and clinical and non-clinical populations.

The symptom scales of the HTQ have been administered in over 74 languages and dialects, and when the instrument was not available for the linguistic and/or cultural group at study, researchers have adapted it to meet their research needs. Although many authors recognized the need to adapt the HTQ, the manner in which they proceeded to make changes to the original usually did not follow the adaptation recommendations set by the developers of this measure, or other recommendations provided by experts in cross-cultural assessment. For instance, only 16% of study authors reported that their adaptation included a translation and some form of cultural adaptation. Of the studies that did not culturally adapt the instrument, none of them reported a cultural verification process that would ensure that the measure was suitable for their sample. Approximately 18% of translations were oral in vivo translations that make it difficult to ensure linguistic and semantic equivalence. Ensuring measurement equivalence by considering both linguistic and cultural elements is critical in cross-cultural assessment (e.g., Flaherty et al., 1988; Keane et al., 1996), and neglecting to take these factors into account can compromise the reliability and validity of scores obtained by the sample in question. Moreover, only one fifth of study authors reported pre-testing the adapted version before commencing data collection. These observations reflect the frequently made observations that (a) the process of translating and cross-culturally validating instruments is usually not considered important in clinical research protocols (Sperber, 2004) and (b) simple forward or back translations without a consensus approach are common in clinical cross-cultural research (Maneesriwongul & Dixon, 2004).

RG Analyses

Although there were 384 studies that used the HTQ symptom scales found in the literature, only 167 of them provided the Cronbach's alpha for their sample. After eliminating

unusable values (i.e., values of duplicate samples) and adding values provided by authors, the final RG analyses included 183 independent reliability estimates coming from 95 studies. With regard to the adaptation procedures reported by the authors, the studies included in the RG analyses appear to be comparable to the overall pool of empirical studies that used the HTQ symptom scales. The results of this RG suggest that the HTQ symptom scales included in the analyses can produce reliable scores across a diverse set of populations. Specifically, according to internal consistency criteria provided by Hunsley and Mash (2008), the mean internal consistency value of the HTQ-30 is excellent (.93) and the mean estimate of the HTQ-16 is good (.89). The unpublished internal consistency values provided by authors had comparable means with the HTQ-30 being rated as "excellent" and the HTQ-16 as "good". If the Spearman-Brown prophecy formula is used to estimate the impact of adding 14 similar items to the HTQ-16 (to compare it to the HTQ-30, which includes the Indochinese culture-related symptoms), the HTQ-16's mean reliability estimate to increases to .94, a slightly higher value than the mean reliability of the HTQ-30. This result suggests that the addition of the cultural items, *per se*, does not add more value to the HTQ-16 in terms of internal consistency. However, this does not imply that the 14 cultural symptoms are irrelevant, as these items may add value to the validity properties of the measure. Furthermore, the majority of studies that used the HTQ-30 did not report culturally adapting the HTQ for their specific sample. These studies thus applied the Indochinese version of the questionnaire to non-Indochinese groups. This implies that the cultural items that are based on clinical findings among Indochinese populations can produce reliable scores across various populations. However, it is unknown whether the addition of other culture-related items could improve the reliability of scores for each of these cultures. Further research is thus needed

to explore the impact of these cultural items on the construct validity of the HTQ across different populations.

Even though the mean reliability estimates of the HTQ-16 and HTQ-30 are good or excellent, the mean reliability estimates for the re-experiencing, avoidance/numbing and arousal subscales are only adequate (.74, .78, .79 respectively). This lower performance in reliability of the subscales may be due to the smaller number of items compared to the total scales (Cortina, 1993). Using the Spearman-Brown formula to evaluate the possible effect on reliability of having 7 items in all subscales increased the re-experiencing and arousal reliability estimates to .83 and .84 respectively. These results indicate that, in principle, the re-experiencing and arousal subscales perform relatively better than the avoidance/numbing subscale in terms of producing reliable scores. This may be due to inconsistent presence of avoidance and numbing symptoms found in trauma studies across various cultures (e.g., Marsella et al., 1996; Norris, Van Landingham, & Vu, 2009) and issues regarding the factorial validity of the 3-factor model of PTSD (e.g. Yufik & Simms, 2010). These issues were highlighted throughout the revision process for the new DSM-5 criteria for PTSD (Hinton & Lewis-Fernández, 2010), during which it was decided to separate the avoidance and numbing symptoms into separate clusters.

Moderator Analyses

The statistically significant *Q* values and high I^2 tests for all scales and subscales suggest that the variability among the reliability estimates is heterogeneous and is not due to sampling error alone. The high degree of variance that is caused by true heterogeneity of the scales indicated that examining sample and methodological characteristics may help explain this variability. However, none of the potential moderators examined accounted for the variance in reliability estimates of the HTQ-16 and HTQ-30. These variables included age, being a sample

for which the HTQ was validated, the type of sampling, and being a forcibly displaced sample. It is possible that there may have been an insufficient number of studies available for the analyses to test these assumptions with sufficient power. Although there may also be other sample or study characteristics that could further explain the variability, it is not immediately obvious what these might be, as the characteristics selected for inclusion in the analyses were those typically shown to impact PTSD severity and/or reliability estimates.

Interestingly, three variables had a moderating effect on the typical reliability of the re-experiencing subscale. One of these moderators was gender, where samples with a higher proportion of women tended to produce more reliable scores. This finding is consistent with another RG study (Vassar, Knaup, Hale, & Hale, 2011) conducted with a different trauma symptoms measure, the Impact of Event Scale (IES; Horowitz, Wilner, & Alvarez, 1979). This RG found a positive correlation between the percentage of female participants and reliability estimates for the intrusion subscale. As symptoms such as flashbacks are intensely emotional in nature, Vassar et al. (2011) posited that women may be more inclined to report experiencing these symptoms due to social norms where women are more considered to be emotionally expressive than men. However, research is needed to determine whether this explanation may account for the present findings with the re-experiencing subscale.

Analyses including cultural group found that reliability coefficients for the re-experiencing subscale were the highest for cultures that were grouped under “African cultures” and the lowest for those grouped under “Asian cultures.” These findings are consistent with suggestions that the experience of symptoms such as flashbacks and nightmares may be related to social norms such as gender-linked expectations for emotional expressivity (Vassar et al., 2011) and cultural valuing of dreams (Shore, Orton, & Manson, 2009). Although these

explanations may account for the present findings with the re-experiencing subscale, research is needed to substantiate such interpretations.

The type of trauma also played a role in the variance of reliability estimates for the re-experiencing subscale. When the types of trauma were grouped into "intentional traumas" and "unintentional traumas," the former had a significantly higher typical reliability estimate than did the latter. As intentional traumas are considered to be more severe these results may indicate that the internal consistency of re-experiencing symptoms of the HTQ is associated with trauma severity. However, to test these interpretations, it would be necessary to have data on trauma severity that was independent of the HTQ scores.

The final significant moderation effect in the analyses was for the cultural orientation of country of origin (individualist vs. collectivist) on the avoidance subscale of the HTQ. Specifically, the reliability coefficient for individualist countries was significantly higher than the typical estimate of collectivist countries. This may be reflective of lower rates of avoidance/numbing symptoms reported in some collectivist cultures, such as Kalahari bushmen (McCall & Resick, 2003) and Vietnamese Americans (Norris et al., 2009). On the other hand, studies including other collectivist cultures have found higher rates of endorsement of avoidance/numbing symptoms (Dyregrov et al., 2000). These variations suggest that more research is needed regarding the role cultural orientation plays in posttraumatic symptoms.

In summary, gender, cultural group and type of trauma had a moderating effect on the reliability of scores of the re-experiencing subscale, and culture orientation had a moderating effect on the reliability of scores of the avoidance/numbing subscale. It is interesting to note that cultural variables (i.e., cultural group and cultural orientation) explained some of the variability for the re-experiencing and avoidance/numbing subscales, but did not for the arousal items. This

pattern offers some support for the perspective that arousal symptoms are consistent across cultures, whereas re-experiencing and avoidance/numbing symptom expression are likely to vary across cultures (e.g., Hinton & Lewis-Fernández, 2011). Further to this point, these results appear to be similar to those found by Rasmussen and colleagues (2015) in a study on the measurement invariance of the HTQ-16. Their results showed that the configural invariance, or overall four-factor structure of the PTSD basic construct as measured by the HTQ-16, was acceptable. However, there were significant response-style differences across cultures, as well as variations in the clinical thresholds of PTSD.

Limitations of the Study

The results of this RG should be interpreted in light of the following limitations. Although scientific standards require that citations be provided for instruments used in a study, the cited reference search strategy that was used may not have identified all published studies that have used the HTQ symptom scales. For the published studies identified with the search strategy, data from many were not included in the RG analyses because Cronbach's alphas were not reported in the primary studies. In fact, only 43.5% of study authors reported a reliability coefficient based on data from their study sample. Even though some authors provided a coefficient alpha after being contacted, there were still many missing reliability estimates that could have been included in the RG. Although there were no significant differences between the unpublished estimates provided by the authors and the published ones, the remaining unpublished estimates may have had an impact on the results. Additionally, there are likely to be studies using the HTQ that have not been published. If these unpublished studies have lower reliability estimates, this would alter the overall mean internal consistency values reported in this study.

In addition, the coding process was difficult at times because the information reported in the studies was often vague and challenging to interpret. There was often a lack of sufficient detail reported about study methods and it is therefore possible that the information reported in many of the studies did not fully represent the actual adaptation procedures that were used. Indeed, insufficient detail in the reporting of adaptation methods appears to be a common problem in the published literature (Maneesriwongul & Dixon, 2004). This issue continues despite reporting standards for research that have been put in place, in part, to ensure the quality of research synthesis and meta-analysis (Appelbaum, Cooper, Maxwell, Stone, & Sher, 2008).

Finally, RG analyses were not conducted on culturally adapted versions of the HTQ due to a limited amount of data. Although there were numerous cultural adaptations found in the literature, there was an insufficient amount of studies for each adapted version of the HTQ that had reliability estimates to conduct meta-analyses. Consequently, the overall psychometric properties of these versions and the adaptation procedures used could not be evaluated.

Implications and Recommendations

As mentioned, using strategies to ensure equivalency is essential in cross-cultural assessment to help minimize bias. Users of the HTQ have access to multiple cross-cultural adaptation guidelines in the literature and the HTQ developers themselves provide recommendations for adapting the instrument for other populations. However, in reality, it appears that researchers may not have followed these adaptation guidelines in their own use of the HTQ, or at least, their specific adaptation procedures have not been reported in published reports. This may be also due to a lack of awareness of adaptation guidelines or reporting standards among researchers, or perhaps journal submission requirements that limit the amount of information provided in the methods section of manuscripts. This, combined with the low rate

in the reporting of reliability estimates, is an indication that more awareness is needed with regards to reporting guidelines for assessment instruments and assessment research. The high cost and lengthy process of adapting instruments can be major challenges, but because the assessment of traumatized individuals can have major clinical and societal implications (e.g., access to treatment, obtaining disability compensation, obtaining refugee status), must take appropriate steps to minimize bias.

The findings of the current meta-analyses indicate that both the HTQ-16 and HTQ-30 are likely to provide reliable scores across diverse populations. However, research results obtained with the subscales should be interpreted cautiously due to lower overall mean reliability estimates and inadequate performance in certain samples. Although these results have important implications for clinicians and researchers as they weigh the strengths and weaknesses of the HTQ, a further evaluation of the construct validity of these scales, specifically convergent and discriminant validity is needed. This, in turn, would provide even further insight into the cross-cultural applicability of this instrument.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing* (5th ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. revised). Washington, DC: Author.

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychological Association Publications and Communications Board Working Group on JARS (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63*, 839-851. doi: 10.1037/0003-066X.63.9.839

Appelbaum, M., Cooper, H, Maxwell, S., Stone, A., & Sher, K. J. (2008). Reporting standards in research in psychology: Why do we need them? What might they be? *American Psychologist, 63*, 839-851. doi: 10.1037/0003-066X.63.9.839

Barnes, L. L. B., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement, 62*, 603-618. doi: 10.1177/0013164402062004005

Barry, A. E., Chaney, B., Piazza-Gardner, A. K., & Chavarria, E. A. (2014). Validity and reliability reporting practices in the field of health education and behavior: a review of

seven journals. *Health Education & Behavior, 41*, 12-18. doi: 10.1177/1090198113483139

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive meta-analysis version 2*. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, UK: John Wiley & Sons, Ltd.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97-111. doi: 10.1002/jrsm.12

Breslau, N., Davis, G., Andreski, P., & Peterson, E. (1991). Traumatic events and posttraumatic stress disorder in an urban population of young adults. *Archives of General Psychiatry, 48*, 216-222. doi:10.1001/archpsyc.1991.01810270028003

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: The 1996 Detroit Area Survey of Trauma. *Archives of General Psychiatry, 55*, 626-632. doi: 10.1001/archpsyc.55.7.626.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322. doi: 10.1111/j.2044-8295.1910.tb00207.x

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101–129. doi: 10.2307/3001666

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46. doi:10.1177/001316446002000104

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications.

Journal of Applied Psychology, *78*, 98-104. doi: 10.1037/0021-9010.78.1.98

Dieckmann, N., Malle, B., & Bodner, T. (2009). An empirical assessment of meta- analytic practice. *Review of General Psychology*, *13*, 101-115. doi:10.1037/a0015107

Dyregrov, A., Gupta, L., Gjestad, R., & Mukanoheli, E. (2000). Trauma exposure and psychological reactions to genocide among Rwandan children. *Journal of Traumatic Stress, 13,* 3–21. doi: 10.1023/A:1007759112499

Elsass, P. (2001). Individual and collective traumatic memories: A qualitative study of posttraumatic stress disorder symptoms in two Latin American localities. *Transcultural Psychiatry, 38,* 306-316. doi: 10.1177/136346150103800302

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

Flaherty, J. A., Gaviria, F. M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J. A., & Birz, S. (1988). Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disease, 176,* 257-263. doi: 10.1097/00005053-198805000-00001

Gagnon, A. J., Tuck, J., & Barkun, L. (2004). A systematic review of questionnaires measuring the health of resettling refugee women. *Health Care for Women International, 25*, 111-149. doi: 10.1080/07399330490267503

Graham, J. M., Diebels, K. J, & Barnow, Z. B. (2011). The reliability of relationship satisfaction: A reliability generalization meta-analysis. *Journal of Family Psychology, 25*, 39-48. doi/10.1037/a0022441

Graham, J. M., Yenling, J. L., & Jeziorski, J. L. (2006). The Dyadic Adjustment Scale: A reliability generalization meta-analysis. *Journal of Marriage and Family, 68,* 701-717.

doi: 10.1111/j.1741-3737.2006.00284.x

Hall, B. W., Ward, A., W. & Comer, C. B. (1988). Published educational research: An empirical study of its quality. *Journal of Educational Research, 81*, 182-189. doi: 10.1080/00220671.1988.10885820

Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2011). *Scientific foundations of clinical assessment.* New York : Routledge.

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, 2*. Retrieved from http://scholarworks.gvsu.edu/cgi/viewcontent.cgi?article=1111&context=orpc

Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement, 61,* 404-420. doi: 10.1177/00131640121971284

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539-1558. doi: 10.1002/sim.1186

Hinton, D. E., & Lewis-Fernández, R. (2010). Idioms of distress among trauma survivors: Subtypes and clinical utility. *Culture, Medicine, and Psychiatry, 34*, 209–218. doi: 10.1007/s11013-010-9175-x

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

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). New York: McGraw-Hill.

Hollifield, M., Warner, T. D., Lian, N., Krakow, B., Jenkins, J. H., Kesler, J., … Westermeyer, J.

(2002). Measuring trauma and health status in refugees: A critical review. *Journal of the American Medical Association, 288,* 611–621. doi: 10.1001/jama.288.5.611

Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41*, 209-218. doi: 10.1097/00006842-197905000-00004

Howell, R. T., & Shields, A. L. (2008). The file drawer problem in reliability generalization: A strategy to compute a Fail-Safe N with reliability coefficients. *Educational and Psychological Measurement, 68*, 120-128. doi: 10.1177/0013164407301528

Hunsley, J., & Mash, E. J. (2008). Developing criteria for evidence-based assessment: An introduction to assessments that work. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 3-14). New York: Oxford University Press.

Jones, E., Vermaas, R. H., McCartney, H., Beech, C., Palmer, I., Hyams, K., & Wessely, S. (2003). Flashbacks and post-traumatic stress disorder: The genesis of a 20th-century diagnosis. *British Journal of Psychiatry, 182,* 158–163. doi: 10.1192/bjp.182.2.158

Keane, T. M., Kaloupek, D. G., & Weathers, F. W. (1996). Ethnocultural considerations in the assessment of PTSD. In A. J. Marsella, M. J. Friedman, E. T. Gerrity, & R. M. Scurfield (Eds.), *Ethnocultural aspects of posttraumatic stress disorder: Issues, research, and clinical applications* (pp. 183-205). Washington, DC: American Psychological Association.

Keane, T. M., Silberbogen, A. K., & Weierich, M. R. (2008). Post-traumatic stress disorder. In J. Hunsley & E.J. Mash (Eds.), *A guide to assessments that work* (pp. 293-315). New York: Oxford University Press.

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry, 52*, 1048-1060. doi: 10.1002/1099-1298(200011/12)

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., Nelson, C. B., & Breslau, N. (1999). Epidemiological risk factors for trauma and PTSD. In R. Yehuda (Ed.), *Risk factors for posttraumatic stress disorder* (pp. 23-60). Washington, DC: American Psychiatric Press.

Kleijn, W. C., Hovens, J. E. & Rodenburg, J. J. (2001). Posttraumatic stress symptoms in refugees: Assessments with the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist-25 in different languages. *Psychological Reports, 88*, 527-532. doi: 10.2466/PR0.88.2.527-532

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., … Kessler, R. C. (2017). Posttraumatic stress disorder in the World Mental Health Surveys, *Psychological Medicine*, 1-15. doi: 10.1017/S0033291717000708

Lange, A., Rietdijk, D., Hudcovicova, M., van de Ven, J. P., Schrieken, B., & Emmelkamp, P. M. (2003). Interapy: A controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. *Journal of Consulting and Clinical Psychology, 71*, 901–909. doi: 10.1037/0022-006X.71.5.901

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal, 333*, 597-600. doi: 10.1136/bmj.333.7568.597

Maneesriwongul, W. & Dixon, J. K. (2004) Instrument translation process: A methods review. *Journal of Advanced Nursing, 48*, 175– 186. doi: 10.1111/j.1365-2648.2004.03185.x

Marsella, A. J., Friedman, M. J., Gerrity, E. T., & Scurfield, R. M. (1996). *Ethnocultural aspects of posttraumatic stress disorder: Issues, research, and clinical applications*. Washington, DC: American Psychological Association.

Matthieu, M., & Ivanoff, A. (2006). Treatment of human-caused trauma: Attrition in the adult outcomes research. *Journal of Interpersonal Violence, 2*, 1654–1664. doi: 10.1177/0886260506294243

McCall, G., & Resick, P. (2003). A pilot study of PTSD symptoms among Kalahari Bushmen. *Journal of Traumatic Stress, 16*, 445–450. doi: 10.1023/A:1025702326392

Mollica, R. F., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., & Lavelle, J. (1992). The Harvard Trauma Questionnaire: Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *The Journal of Nervous and Mental Disease, 180,* 111-116. doi: 10.1097/00005053-199202000-00008

Mollica, R. F., Caspi-Yavin, Y., Lavelle, J., Tor, S., Yang, T., Chan, S., ... de Marneffe, D. (1996). The Harvard Trauma (HTQ) Manual: Cambodian, Laotian, and Vietnamese versions. *Torture, 6*(1), 19-33.

Nakeyar, C., & Frewen, P. A. (2016). Evidence-based care for Iraqi, Kurdish, and Syrian asylum seekers and refugees of the Syrian civil war: A systematic review. *Canadian Psychology, 57*, 233-245. doi: 10.1037/cap0000067

Norris, F. H. (1992). Epidemiology of trauma: frequency and impact of different potentially traumatic events on different demographic groups. *Journal of Consulting and Clinical Psychology, 60,* 409-418. doi: 10.1037/0022-006X.60.3.409

Norris, F. H., Van Landingham, M. J., & Vu, L. (2009). PTSD in Vietnamese Americans following Hurricane Katrina: Prevalence, patterns, and predictors. *Journal of Traumatic*

Stress, 22, 91-101. doi: 10.1002/jts.20389

Perkonigg, A., Kessler, R. C., Storz, S., & Wittchen, H.-U. (2000). Traumatic events and post-traumatic stress disorder in the community: Prevalence, risk factors and comorbidity. *Acta Psychiatrica Scandinavica, 101*, 46-59. doi: 10.1034/j.1600-0447.2000.101001046.x

Pietrzak, R. H., Goldstein, R. B., Southwick, S. M., & Grant, B. F. (2011). Prevalence and Axis I comorbidity of full and partial posttraumatic stress disorder in the United States: Results from Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Anxiety Disorders, 25,* 456-465. doi: 10.1016/j.janxdis.2010.11.010

Pole, N., Gone, J. P., & Kulkarni, M (2008). Posttraumatic stress disorder among ethnoracial minorities in the United States. *Clinical Psychology: Science and Practice, 15*, 35–61. doi: 10.1111/j.1468-2850.2008.00109.x

Porter, M., & Haslam, N. (2005). Predisplacement and postdisplacement factors associated with mental health of refugees and internally displaced persons: A meta-analysis. *Journal of the American Medical Association, 294*, 602-612. doi: 10.1001/jama.294.5.602.

Qualls, A. L. & Moss, A. D. (1996). The degree of congruence between test standards and test documentation within journal publications. *Educational and Psychological Measurement, 56,* 209-214. doi: 10.1177/0013164496056002002

Rasmussen, A., Verkuilen, J., Ho, E., & Fan, Y. (2015). Posttraumatic stress disorder among refugees: Measurement invariance of Harvard Trauma Questionnaire scores across global regions and response patterns. *Psychological Assessment, 27*, 1160-1170. doi: 10.1037/pas0000115

Renner, W., Salem, I., & Ottomeyer, K. (2007). Posttraumatic stress in asylum seekers from Chechnya, Afghanistan, and West Africa: Differential findings obtained by quantitative and qualitative methods in three Austrian samples. In J. P. Wilson & C. C. Tang (Eds.), *Cross-cultural assessment of psychological trauma and PTSD* (pp. 239-278). Boston, MA: Springer.

Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 11*, 306-322. doi:10.1037/1082-989X.11.3.306

Shoeb, M., Weinstein, H., & Mollica, R. (2007). The Harvard Trauma Questionnaire: Adapting a cross-cultural instrument for measuring torture, trauma and posttraumatic stress disorder in Iraqi refugees. *The International Journal of Social Psychiatry, 53*, 447-463. doi: 10.1177/0020764007078362

Shore, J. H., Orton, H., & Manson, S. M. (2009). Trauma-related nightmares among American Indian veterans: Views from the dream catcher. *American Indian and Alaska Native Mental Health Research, 16*, 25-38. doi : 10.5820/aian.1601.2009.25

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428. doi: 10.1037/0033-2909.86.2.420

Spearman, C. C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271-295. doi: 10.1111/j.2044-8295.1910.tb00206.x

Sperber, A. D. (2004) Translation and validation of study instruments for cross-cultural research. *Gastroenterology, 126*, S124-S128. doi: 10.1053/j.gastro.2003.10.016

Steel, Z., Chey, T., Silove, D., Marmane, C., Bryant, R. A., & van Ommeren, M. (2009). Association of torture and other potentially traumatic events with mental health outcomes among populations exposed to mass conflict and displacement: A systematic review and

meta-analysis. *Journal of the American Medical Association, 302,* 537-549. doi: 10.1001/jama.2009.1132

Therrien, Z., & Hunsley, J. (2013). Assessment of anxiety in older adults: A reliability generalization meta-analysis of commonly used measures. *Clinical Gerontologist 36,* 171-194. doi:10.1080/07317115.2013.767871

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54*(4), 837-847.

Thompson, B. & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60,* 174-195. doi: 10.1177/0013164400602002

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58,* 6-20. doi: 10.1177/0013164498058001002

Vacha-Haase, T., Kogan, L.R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement, 60,* 509–522. doi: 10.1177/00131640021970682

Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development, 44,* 159-168. doi: 10.1177/0748175611409845

Van der Velden, P. G., Grievink, L., Kleber, R. J., Drogendijk, A. N., Roskam, A. J., Marcelissen, F. G., … Gersons, B. P. (2006). Post-disaster mental health problems and the utilization of mental health services: A four-year longitudinal comparative study.

Administration and Policy in Mental Health and Mental Health Services Research, *33*, 279-288. doi: 10.1007/s10488-005-0027-x

Vassar, M., Knaup, K.G., Hale, W., & Hale, H. (2011). A meta-analysis of coefficient alpha for the Impact of Event Scales: A reliability generalization study. *South African Journal of Psychology*, *41*, 6-16. doi: 10.1177/008124631104100102

Wood, J. (2008). Methodology for dealing with duplicate study effects in meta-analysis. *Organizational Research Methods, 11*, 79-95. doi: 10.1177/1094428106296638

Yufik, T., & Simms, L. J. (2010). A Meta-analytic investigation of the structure of posttraumatic stress symptoms. *Journal of Abnormal Psychology, 119*, 764-776. doi: 10.1037/a0020981

Figure 1

Literature Search Flow Chart

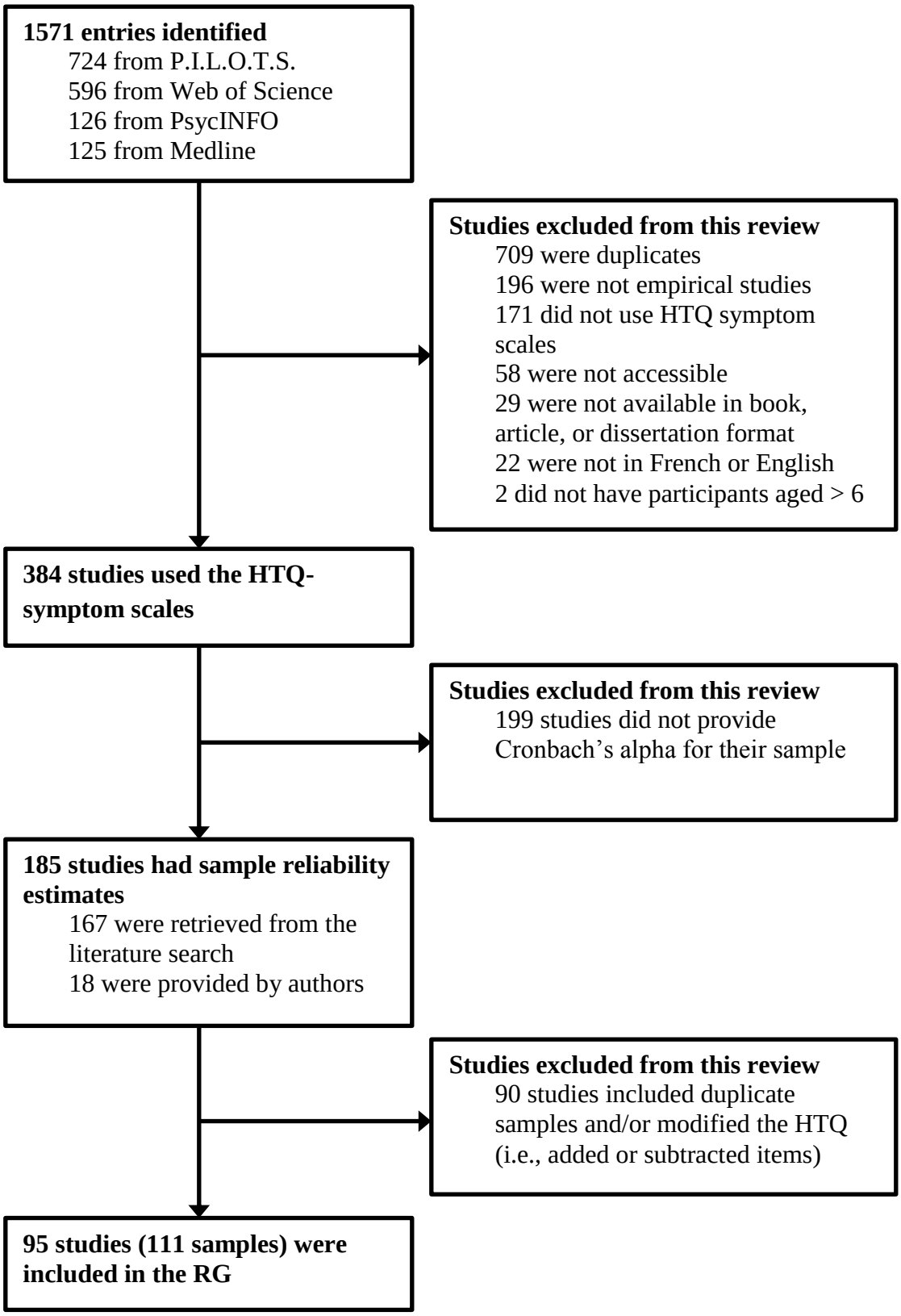

Table 1

Descriptive Statistics of the Adaptation and Administration Procedures Reported in Samples Using the HTQ Symptom Scales

	Overall pool of samples	Samples included in the RG analyses
	n (%)	*n* (%)
Total number of samples	242	111
Required an adaptation	199 (82.2)	97 (87.4)
Total adaptations of the HTQ	181 (74.8)	85 (76.6)
Reported adaptation methods	127	57
Translation only	95 (74.8)	42 (73.7.4)
Cultural adaptation only	3 (2.4)	1 (1.8)
Translation and cultural adaptation	29 (16.0)	14 (24.6)
Used a consensus approach	60 (47.2)	30 (52.6)
Conducted a pre-test	28 (22.0)	7 (12.3)
Used HTQ adaptation guidelines	12 (9.4)	9 (15.8)
Used other adaptation guidelines	30 (23.6)	14 (24.6)
Reported translation procedures	124	56
In vivo oral translation	22 (17.7)	8 (14.3)
Forward translation	25 (20.2)	13 (23.2)
Back translation	54 (43.5)	28 (50)
Blind back translation	23 (18.5)	7 (12.5)
Reported administration procedures	226	108
Self-report	105 (46.5)	61 (56.5)
Non-clinician administered	50 (22.1)	20 (18.5)
Clinician administered	67 (29.6)	23 (21.3)
Mixed methods	4 (1.8)	4 (3.7)

Table 2

Descriptive Statistics of Studies Included in the RG

Scale	Items	Total sample size	Mean sample size	Mean age (years)	Gender composition (% males)
HTQ-30	30	9,322	259	37.14	44.04
HTQ-16	16	23,721	339	35.47	50.62
Re-experiencing	4	6,731	306	36.69	38.28
Avoidance/numbing	7	7,910	293	35.01	38.05
Arousal	5	8,094	289	34.87	38.13

Table 3

Descriptive Statistics for the Reliability Values of the HTQ Symptom Scales and Subscales

Scale	*k*	Mean α	*95% confidence interval* Lower	Upper	Min.	Max.	*Q*	I^2	Fail-Safe *N* for RG
HTQ-30	36	.93	.92	.94	.86	.98	684.27*	94.89	339
HTQ-16	70	.89	.87	.90	.73	.98	1786.34*	96.14	272
Re-experiencing	22	.74	.72	.77	.43	.84	100.08*	79.02	77
Avoidance/ numbing	27	.78	.75	.81	.54	.91	258.93*	89.96	222
Arousal	28	.79	.76	.82	.57	.90	333.65*	91.91	198

Note. Min. = Minimum reliability estimate reported in the literature; Max. = Maximum reliability estimate reported in the literature; *k* = number of samples included in the analyses; *p < .001

Table 4

Summary of Moderating Effects for the Internal Consistency of the HTQ Symptom Scales

Moderator	HTQ-30 *N* = 36	HTQ-16 *N* = 70	Re-experiencing *N* = 22	Avoidance *N* = 27	Arousal *N* = 28
Age	No (28*)	No (52)	No (21)	No (25)	No (26)
Gender	No (34)	No (62)	Yes (21)	No (26)	No (27)
Cultural group	No (34)	No (59)	Yes (22)	No (27)	No (28)
Cultural orientation	No (21)	No (30)	No (21)	Yes (23)	No (24)
Forcibly displaced sample	No (31)	No (68)	No (22)	No (27)	No (28)
Recruitment site	No (36)	No (67)	No (22)	No (27)	No (28)
Trauma type	No (32)	No (64)	Yes (18)	No (20)	No (21)
Original sample	No (36)	No (70)	N/A	N/A	N/A

Note. *Total samples included in the moderator analyses are in parentheses; Yes = Significant moderating effect; No = No significant moderating effect; N/A = Insufficient information available to conduct moderator analyses

Study 2

A Validity Generalization Study of the Harvard Trauma Questionnaire

Abstract

The current study examined the construct validity of the symptom scales of the Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992), a commonly used cross-cultural instrument that aims to assess the psychological effects of trauma. A cited-reference search was conducted to locate all publications and dissertation that used the symptoms scales of the HTQ. Random-effects validity generalization (VG) meta-analyses were performed on discriminant and convergent validity coefficients. These analyses included 125 validity coefficients from 78 studies, representing data from 21,156 individuals. The variance of score validity was characterized by considering sample and methodological variables that are predictive of the validity coefficients. The findings of the VG analyses indicate that the convergent validity properties of the HTQ-16 are supported to some extent, and that this scale performs rather questionably in terms of discriminant validity. Furthermore, there is limited support for either the convergent or discriminant validity of the HTQ-30. Despite some encouraging results, more work is needed to establish the validity of these scales. We discuss the complexities involved in establishing the construct validity PTSD assessments in light of the high comorbidity rates and over lapping symptoms with other disorders.

Keywords: posttraumatic stress disorder, assessment, cross-cultural assessment, construct validity, meta-analysis

A Validity Generalization Study of the Harvard Trauma Questionnaire

Assessment instruments rely heavily on the conceptualization of the construct they are intended to measure. If the conceptual underpinnings of the construct are not solid, then the psychometric properties of the instrument can be seriously compromised (Haynes, Smith, & Hunsley, 2011). One psychological construct that has been extensively debated in regards to its conceptualization is posttraumatic stress disorder (PTSD). A central concern is that the PTSD construct was developed according to Western cultures' conceptualizations of trauma and its sequelae, and may not be applicable to other cultures (e.g., Jones et al., 2003). Consequently, it is not surprising that there has also been much questioning regarding the cross-cultural applicability of screening tools used to assess PTSD. The objective of the current study is to shed some light on this issue by examining the construct validity of the Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992), a PTSD measure that was designed to be adapted and used across cultures.

The Harvard Trauma Questionnaire

The HTQ was developed by Mollica and colleagues (1992) and has been validated by its developers for various cultural and linguistic groups (Shoeb, Weinstein, & Mollica, 2007).Most versions of the HTQ consist of four parts. Part 1 measures different traumatic events, part 2 is an open-ended description of the most traumatic event experienced by the respondent, part 3 evaluates the circumstances surrounding possible head injury and, finally, part 4 is a list of trauma symptoms. The first 16 items in part 4 are derived from the DSM-III-R/IV (American Psychiatric Association, 1987, 1994) PTSD criteria, and they are the same in every version of the HTQ. The second set of trauma symptom items in part 4 are culture-specific questions and are tailored for each version of the HTQ. The HTQ has also been translated and adapted in many other languages by other researchers (e.g. Kleijn, Hovens, & Rodenburg, 2001). Systematic

reviews of tools used to assess the health of refugees have found that the HTQ has been extensively used by researchers to assess trauma and its sequelae (Gagnon, Tuck, & Barkun, 2004; Hollifield et al., 2002), and has been shown to be useful to exclude PTSD non-cases in certain forcibly displaced populations (Nakeyar & Frewen, 2016). In addition, in their review of PTSD measures, Keane, Silberbogen, and Weierich (2008) recommended the use of the HTQ for the assessment of PTSD across cultures. They described the HTQ as having "linguistic equivalence across the many cultures and languages with which it has been used thus far" (p. 297). However, they did not provide sufficient empirical evidence to support the cross-cultural applicability of the instrument, or evaluate it according to established evidence-based assessment guidelines, such as Hunsley and Mash's (2008).

We recently conducted a reliability generalization (RG) study to evaluate the internal consistency of HTQ scores through meta-analysis (see Darzi & Hunsley, 2017). The results of this RG showed that the HTQ-16 and HTQ-30 are likely to provide reliable scores across a large diversity of populations. However, the re-experiencing, avoidance/numbing and arousal subscales should be used with caution due to lower overall mean reliability estimates and inadequate performance in certain samples. With regards to the validity of the HTQ symptom scale, some authors have suggested that it may not be generalizable due to having been developed from a psychiatric outpatient population (Hollifield et al., 2002). In a more recent study, Rasmussen and colleagues (2015) found that the configural invariance, or overall four-factor structure of the PTSD construct as measured by the HTQ-16, was acceptable among individuals of various cultural groups. However, there were significant response-style differences across cultures, as well as variations in the clinical thresholds of PTSD. Another study on the factor structure of PTSD using the HTQ-16 compared three models across three groups living in

non-Western low- and middle-income countries (Michalopoulos,, et al., 2015). The results showed that all models had an adequate fit for two cultural groups (i.e. Congolese and Burmese samples), however none of the models had an adequate fit for the other (i.e. Northern Iraqi sample). These important findings speak to the potential limitations of the construct validity of the HTQ symptom scales.

Although these studies provide preliminary evidence of the validity of scores provided by the HTQ, its overall construct validity properties with different cultural groups remains unknown. Because this instrument is intended to be used across cultures, a further analysis of empirical studies that have used the HTQ in different populations is warranted. Aggregating empirical findings across studies and cultures will provide pertinent information regarding the potential cross-cultural utility of the HTQ, thereby helping clinicians and researchers decide whether it is an appropriate assessment tool for use with individuals of different cultural backgrounds.

Validity and Validity Generalization

Validity refers to the extent to which an assessment instrument measures what it is supposed to measure and is considered the most fundamental element of test development and evaluation (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). An instrument that has been evaluated as having good validity has variation in scores that reflects the true variation in the construct that is being measured (Haynes et al., 2011).

Although there are various kinds of validity, the overarching form of validity is construct validity. This type refers to the degree to which a test score reflects the construct of interest, and includes both convergent and discriminant validity. Evidence for convergent validity refers to the

extent to which the scores of an instrument are related to scores of a different measure that is intended to assess the same construct, or other variables that have theoretically established associations with the construct (Messick, 1995). Evidence for discriminant validity refers to the degree to which the scores of an instrument are unrelated to measures of unrelated constructs (Haynes et al., 2011). Thus, low correlations between the two instruments can be indicative of good discriminant validity.

Validity is not a static characteristic of an instrument but can vary according to context and different sample characteristics. This was described as "situational specificity" by Schmidt and Hunter (1998) and was used to support their hypothesis that there are statistical artifacts (e.g., sampling error, measurement error) that can cause instruments to demonstrate adequate evidence of validity in one situation and inadequate evidence of validity in another. This mirrors the concept that reliability estimates are sample-dependent. Those using the instrument are encouraged to examine validity in their specific population and purpose in using the instrument (Kelly, O'Malley, Kallen, & Ford, 2005).

Establishing the validity of scores produced by an instrument is important as it also addresses the validity of theories on the construct being measured and the operationalization of the construct (Haynes et al., 2011; Kimberlin & Winterstein, 2008). The evidence of validity of a measure continuously evolves as the theories of the constructs being measured change according to new research developments (Haynes et al., 2011). Thus, evaluating the validity of measures should be an ongoing process. Due to the controversy regarding the cross-cultural applicability of the PTSD construct, and the abundance of new research regarding posttraumatic reactions across the globe, the evaluation of instruments that are intended to measure these symptoms across cultures, such as the HTQ, is of utmost importance.

Following their initial observations regarding situation specificity in validity, Schmidt and Hunter (1977) developed a meta-analytic method called "validity generalization" (VG) to examine the correlations between a specific test and a criterion. Specifically, VG estimates a mean validity coefficient of a measure based on the correlations between the measure and validity criteria of different studies. For example, to evaluate the discriminant validity properties of a PTSD questionnaire, one could gather relevant studies that provide correlations (or other measures of the relations between two variables) between this questionnaire and a measure of impression management, and then calculate a mean correlation. This method can also provide indications of the effects that might account for variability in the values found in a given sample.

The VG method has been primarily used in the field of employment selection and testing (e.g., Dye, Reck, & McDaniel, 1993; Ones, Viswesvaran, & Schmidt, 2003) and has been recommended in the Standards for Educational and Psychological Testing as a general approach to evaluate validity (AERA et al., 2014). According to these standards, VG should preferably be used when certain conditions are met: when there is a large meta-analytic database, when these data represent more or less the type of situation to which the use of scores on the instrument will be generalized, and when statistical artefacts are corrected and this correction produces validity evidence that is consistent (AERA et al., 2014).

The Current Study

The present research was designed to evaluate the extent to which the HTQ symptom scales have produced valid scores across studies found in the literature. Specifically, the convergent and discriminant validity of these scales across samples were investigated by using validity generalization analyses. This meta-analytic method determined the mean convergent and discriminant validity estimates obtained on the HTQ symptom scales and examined the

variability in score validity across samples through moderator analyses. Specifically, moderator analyses were conducted to assess the potential impact of sample and methodological characteristics on scores produced by the HTQ symptom scales and subscales, thus further examining the potential sources of variance. Similar to Study 1, variables that have shown mean differences between groups with and without PTSD were examined.

Six sample characteristics were included as potential moderators. First, age was considered a relevant potential source of variance as some studies suggest that the rates of developing PTSD after exposure to a traumatic event may decline as age increases (Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995; Norris, 1992). Second, gender was also included as epidemiological research shows that compared to men, women tend to experience fewer traumatic events but are more likely to develop PTSD (Breslau et al., 1998; Perkonigg, Kessler, Storz, & Wittchen, 2000).

Because this study was designed to evaluate the cross-cultural applicability of the HTQ the impact of the cultural background of study participants of the study was included as a third potential moderator. In extension to this, the cultural background of the sample in terms of the constructs of collectivism versus individualism was examined for its possible moderating effects. Such differences in perspective may have an impact on symptom reporting, and consequently, on score validity. To categorize cultures as collectivist or individualist in the current study, Hofstede's dimensions of national culture were used (Hofstede, Hofstede, & Minkov, 2010). As a result, only the countries in Hofstede's research studies were categorized and included in the moderator analyses.

Another sample characteristic that was included as a possible moderator was the type of trauma that was separated into two categories: intentional trauma (e.g., war, torture, sexual

assault) and unintentional trauma (e.g., natural disasters, accidents). These categories were chosen because studies have shown that individuals who are exposed to intentional traumas had worse health outcomes than did individuals who experienced traumas that were not deliberate (Lange et al., 2003; Matthieu & Ivanoff, 2006; Van der Velden et al., 2006). Whether study participants were forcibly displaced was also included as a sixth moderator variable, as researchers tend to find higher rates of psychological difficulties among individuals who have been displaced as the result of war as compared to non-displaced groups (Fazel, Wheeler, & Danesh, 2005; Porter & Haslam, 2005; Steel et al., 2009).

Two methodological variables were included as potential moderators. The sites from which the samples were recruited were examined. Psychometric properties of instruments can be dependent on the severity of the disorder measured (Haynes et al., 2011) and, as symptoms are generally more severe in individuals who are seeking psychotherapy or medical treatment than they are in individuals from non-clinical samples, the effect of recruit site (i.e., clinical, community, or mixed clinical and community) was considered. Convergent validity estimates should be highest and discriminant validity estimates should be lowest in studies involving samples for which the HTQ was originally intended to be used. Accordingly, moderator analyses were conducted to determine whether having a study sample that was culturally similar to the samples used in the development of the HTQ affects validity estimates.

Method

Literature Search and Data Collection

The details of the search strategy to identify potential studies to be included in the VG and inclusion criteria used are reported elsewhere (Darzi & Hunsley, 2017). A flow chart of the data reduction process can be seen in Figure 1.

To ensure the statistical independence of the studies to be included in the analyses, multiple publications on the same data set were identified. The detection heuristic developed by Wood (2008) was used as a guideline to help identify same samples. This strategy included determining whether the study and sample characteristic were similar in the suspected multiple publications and verifying if the same authors were included in the studies. To help decide which study to eliminate, several factors were considered, such as the amount of descriptive data reported and the reporting of validity coefficients. Furthermore, when studies provided validity coefficients of different samples within the same study, each sample was identified as separate and retained for the analyses.

In summary, there was a total of 384 studies that matched the inclusion criteria and a total of 116 studies that had convergent and/or discriminant validity estimates. After removing studies using the same samples, studies that included modified versions of the HTQ symptom scales (i.e., added or removed items), and unusable validity information (such as a lack of information necessary to compute a validity coefficient), there were a final total of 78 studies and 79 samples to be included in the analyses.

Coding of descriptive data and moderators. The relevant articles were coded for convergent and discriminant validity information (i.e., construct, measure, format, values), sample characteristics, and methodological characteristics for the HTQ symptom scales. The initial coding was performed by the first author (a clinical psychology doctoral student with previous coding experience) and was facilitated by a coding manual (see Appendix A) that was based on a review of the literature on trauma, cross-cultural measurement and validity generalization. The data were entered on a coding sheet (see Appendix B). The study and sample variables were used in the primary descriptive statistics of the overall pool of studies

included in the meta-analyses. Some variables were also examined as potential moderators. Study and methodological characteristics included the following categorical variables: (a) language of administration of the HTQ; (b) country of study; (c) type of sampling (community; seeking mental health services; seeking medical treatment sample); (d) version of the HTQ (original or adapted version), (e) whether an adaptation was needed for the sample (yes or no); (f) adaptation procedure used (cultural adaptation only, translation only, or both); (g) adaptation procedure followed the developers' recommendations (yes or no), (h) adaptation procedure followed other experts' recommendations (yes or no); (i) translation procedure used (oral translation; forward translation; back translation; blind back translation); (j) a consensus approach was used during adaptation (yes or no); (k) the adaptation included a pre-test (yes or no); and (l) type of administration (clinician administered; non-clinician administered; self-report).

Sample characteristics included the following continuous variables: sample size, the mean age of participants, and gender composition of the sample (percentage of males). Sample characteristics also included the following categorical variables: (a) country of origin of participants; (b) country of origin's cultural orientation (individualist or collectivist) based on Hofstede's categorization (only countries classified by Hofstede were coded); (c) type of trauma experienced (intentional or unintentional); and (d) sample consisted of forcibly displaced persons (yes or no).

The second author (a clinical psychologist and professor with extensive experience in meta-analysis) coded 20% of the studies that were included in the meta-analyses to assess the reliability of the coding procedures. The inter-coder reliability analyses were performed using SPSS version 20. The kappa (*k*) statistic was used for the categorical variables and intraclass

coefficients (ICC) were used to assess the inter-coder reliability of continuous variables. These analyses were important to help eliminate random variation and increase the reliability and power of the results (Dieckmann, Malle, & Bodner, 2009). The inter-coder reliability had a rating of "good" as per Hunsley and Mash's (2008) criteria with a mean $k = 0.77$ and ICC coefficient of 0.85. Several discrepancies between coders were caused by unclear reporting of data in the studies, and vague descriptions of the methodological and adaptation procedures. Discrepancies were discussed until a unanimous agreement between the coders was reached.

Data Analysis

Calculating mean effect sizes. Several steps were followed to develop convergent and discriminant validity coefficient estimates for the HTQ-16 and HTQ-30s. The first step was to enter the validity information of each sample (sample identifier, sample size, and validity coefficient) into the Comprehensive Meta-Analysis software version 2.2.064 (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005). Values of indexes that reported the variance shared between a HTQ symptom scale and a convergent or discriminant validity construct (e.g., *r*, *d*) were included in the analyses. Effect sizes that were not reported as correlations were converted into *r* with the CMA software. A Fisher's *r*-to-*z* transformation was then performed to compensate for the high level of skewness and non-normal distribution usually found with correlations. All further analyses were performed with these transformations (see Borenstein, Hedges, Higgins, & Rothstein, 2009).

A relative weight and a 95% confidence interval were calculated for each sample. This relative weight, or inverse-variance, is primarily based on the sample size (i.e., smaller sample sizes are generally assigned less weight). An overall mean effect size, its statistical significance and upper and lower 95% confidence intervals were then calculated. The Fisher's *z* values were

then converted back to their original *r* values to facilitate interpretation.

Random-effects model. A random-effects model was chosen to allow for the true effect to be different across samples. This model is in contrast to the fixed-effect model that assumes that the samples have a common true effect and that the differences in observed effects are due entirely to sampling error alone (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2010). Because the random-effects model considers that the effect sizes may come from different heterogeneous populations, this model was considered the most suitable to include in the VG analyses which were comprised of highly heterogeneous samples.

Heterogeneity of effects. A *Q* test of homogeneity (or Cochran's *Q*; Cochran, 1954) and I^2 index (Higgins & Thompson, 2002) were used to evaluate the degree of dispersion of validity estimates around the mean validity coefficients. The *Q* statistic evaluates the null hypothesis that the studies included in the meta-analysis have a common effect size and a significant *Q* statistic suggests that moderator analyses should be conducted to help find the sources of heterogeneity. The I^2 index assesses the proportion of the observed variance that is caused by true heterogeneity rather than being due to chance alone (Higgins & Thompson, 2002). The following benchmarks proposed by Higgins and Thompson (2002) were used to assess the different levels of heterogeneity and interpret the I^2 index: 25% = low, 50% = moderate, and 75% = high.

Possibility of publication bias. To examine the possible influence that unpublished effects may have on the mean validity values obtained in the meta-analyses, an Orwin's fail-safe *N* (Orwin, 1983) was calculated for each meta-analysis. The result of this procedure is an estimate of the number of studies with estimate of .01 would be needed lower the average validity estimate to under .2 (which is the cut-off for a small effect). Funnel plots were not used

to assess publication bias because of the risk of visual misinterpretation (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006).

Analysis of moderators. Moderator analyses were performed to examine possible sources of variance on estimates that had statistically significant heterogeneity. Mixed effects (method of moments) meta-regression analyses were performed to examine the relations between continuous variables and score validity. For categorical variables, a series of random-effects subgroup analyses were used to examine the relations between the validity coefficients and study and sample characteristics. Because some of the variables had an insufficient number of studies per subgroup (i.e., less than 2) some categories were merged together to help increase the power of the analyses. For instance, "culture" was separated according to larger geographical cultural groups according to continent of origin (e.g., Asian, European, African), and "types of trauma" were categorized as "intentional trauma" and "unintentional trauma".

Results

Overview of Studies included in the Meta-Analyses and Descriptive Statistics of Potential Moderators

The present meta-analyses included 79 samples from 78 studies, representing data from 21,156 individuals (see Appendix D for the list of studies included). The mean sample size was 280, the mean age of participants was 37 years, and, on average, 46.6% of study participants were male. The studies were conducted in over 36 countries, with samples coming from over 35 different countries. Consistent with this, the HTQ symptom scales were administered in over 42 different languages and dialects.

Of the study samples that reported the sampling procedure (n = 79), 67.1% recruited participants from the community, 22.8% recruited participants who were seeking mental health

services, 6.3% recruited participants who were seeking medical treatment, and 3.8% recruited from multiple sources. Fifty-nine percent of the samples were composed of forcibly displaced individuals (i.e., refugees, asylum seekers, internally displaced people) as opposed to natives or immigrants. The most common types of traumatic events were intentional traumas, with 79.7% of samples having all or most individuals who experienced these traumas. Unintentional traumas were experienced by 12.7% of the samples and the remainder 7.6% of samples consisted of individuals who had been exposed to both types of traumas or none.

In terms of how the HTQ symptom scales were used within the independent samples, 74.7% (n = 59) involved adaptations, although 86.1% (n = 68) required the use of an adapted version of the measure because of the nature of the samples. Of those study samples that used adaptations, 66.1% were translations only, 10.2% were both translation and cultural adaptation, and 23.7% made no mention of how they adapted the measure. Of the studies that did not culturally adapt the instrument, none of them reported a cultural verification process that would ensure that the measure was suitable for their sample. Regarding translation procedures, of those that reported translating the instrument (n = 45), 20% indicated using a forward translation, 24.4% used in vivo oral translations, 35.6% used a back translation approach, and 20% used a blind back translation approach. Among the studies that reported their adaptation procedures (n = 45), 6.7% reported using the adaptation recommendations provided by the developers of the HTQ and 26.7% reported following the recommendations from other authorities on measure adaptation. With regards to steps taken to ensure the validity of the translation or adaptation, 37.8% used a consensus approach and 22.2% conducted a pre-test. Finally, of those who reported the administration procedures used (n = 75), 40% were administered via self-report, 24% were administrations done by non-clinicians, and 37.3% were clinician-administered.

Validity Generalization Analyses

Convergent validity. Altogether, there were 72 convergent validity coefficients entered in the meta-analyses. When a study reported two or more correlations for the same construct, they were averaged to ensure statistical independence. Consistent with procedures used in other VG studies (e.g., Kinicki, McKee-Ryan, Schriescheim, & Carson, 2002), the various correlates were grouped into broad correlate categories. These categories included the following constructs: (a) cumulative trauma (e.g., number of different trauma types experienced as per the HTQ events scale); (b) other posttraumatic stress measure (e.g., Impact of Event Scale [IES; Horowitz, Wilner, & Alvarez, 1979); (c) cultural expressions of trauma (e.g., additional cultural items of the HTQ symptom scales); and (d) dissociation measure (e.g., dissociation subscale of the Trauma Symptom Checklist [TSC; Briere &Runtz, 1989]). Descriptive statistics and results of the convergent VG of each of these categories are presented in Table 1. The mean convergent validity coefficients for the HTQ-16 ranged from $r = .32$, $p < 0.001$, 95% CI [0.26, 0.38] for the "cumulative trauma" category to $r = .75$, $p < 0.001$, 95% CI [0.59, 0.86] for "cultural expressions of trauma." For the HTQ-30, a meta-analysis could only be conducted with the "cumulative trauma" category because of the insufficient number of correlations available for the other constructs. The mean convergent validity estimate for this category was $r = .26$, $p < 0.001$, 95% CI [0.17, 0.34].

All convergent VG analyses had a significant *Q* statistic. The significant *Q* statistics and high I^2 suggest that there is an important amount of variance across samples for those specific VG results, therefore further examination of potential study and sample characteristics that account for this variability was warranted.

Discriminant validity. A total of 53 discriminant validity coefficients were entered in the meta-analyses examining discriminant validity. The various correlates were separated into the following correlate categories: (a) depression and anxiety symptoms measures (e.g., Hopkins Symptom Checklist-25 [HSCL-25: Mollica, Wyshak, de Marneff, & Lavelle, 1987]); (b) anxiety symptoms measures (e.g., anxiety subscale of the General Health Questionnaire (GHQ; Goldberg & Williams, 1988); (c) depression symptoms measures (e.g., Beck Depression Inventory-II [BDI; Beck, Steer, & Brown, 1996]); and (d) somatization measures (e.g., HSCL somatization scale). Descriptive statistics and results of the discriminant VG of each of these categories are presented in Table 2. The mean discriminant validity coefficients for the HTQ-16 ranged from r = .43, p < 0.001, 95% CI [0.28, 0.56] for the "somatization measures" category to r = .80, p < 0.001, 95% CI [0.77, 0.83] for "depression and anxiety symptoms measures." The mean discriminant validity coefficients for the HTQ-30 ranged from r = .45, p = 0.239, 95% CI [-0.31, 0.86] for the "somatization measures" category to r = .79, p < 0.001, 95% CI [0.61, 0.89] for "depressive and anxiety symptoms measures."

The "depression and anxiety symptoms measures" category for the HTQ-16 and the "depression symptoms measures" category for the HTQ-30 had nonsignificant Q statistics. The significant Q statistics and high I^2 of the other discriminant VG analyses indicated a need to further examine the role of potential moderators.

Publication bias. Because the CMA software cannot compute an Orwin's fail-safe N for meta-analyses with fewer than 3 studies, three of the VG studies could not be tested for publication bias. A generally accepted interpretation of fail-safe N is that a meta-analytic result is unlikely to be affected by publication bias if the calculated $N \geq 5k + 10$ (Orwin, 1983). Based on this formula, none of the VG results in this study are robust to publication bias, and should be

considered as preliminary results until additional data on these categories of validity coefficients are available.

Moderator analyses. A summary of the moderating effects is presented in Table 3. Samples that had missing information for a specific variable were excluded from these analyses. Meta-regression analyses (mixed effects method of moments) were conducted to assess the potential impact of age (mean age of sample) and gender (percentage of participants who were males). Age was a significant moderator for the association between the "somatization measure" category and HTQ-16, β = -0.092, Q_{model} = 4.60, $p < 0.05$, 95% CI [0.00, 0.03]. Specifically, higher associations between the HTQ-16 and somatization measures were found with lower mean ages of study samples. Gender was also a moderator for this association, β = -0.014, Q_{model} = 10.49, $p < 0.05$, 95% CI [-0.02, -0.01], with study samples comprised of mainly female participants having higher validity estimates.

With regards to the subgroup analyses, the "forcibly displaced sample" and "country of origin's cultural orientation" variables were not significant moderators for any VG result. The "cultural group" variable was significant for the convergent association between the "cumulative trauma" variable and the HTQ-30. Specifically, "African cultures" had the lowest mean validity coefficient (r = 0.06, p = 0.100, 95% CI [-0.01, 0.13]) and "Asian cultures" had the highest mean validity coefficient (r = 0.40, $p < 0.001$, 95% CI [0.32, 0.48]). The summary effect sizes for "cultural group" were also significantly different from each other for the discriminant association between the "somatization measures" variable and the HTQ-16. In this case, "European culture" had the lowest coefficient (r = 0.17, p = 0.088, 95% CI [-0.03, 0.35]) and "Asian culture" the highest (r = 0.60, $p < 0.001$, 95% CI [0.56, 0.64]). The "recruitment setting" variable had a significant effect for two convergent VG results: (a) the association between "cumulative

trauma" and the HTQ-30, where the lowest mean coefficient was for "medical setting" (r = 0.04, p = 0.356, 95% CI [-0.05, 0.12]) and the highest for "community setting" (r = 0.33, p < 0.001, 95% CI [0.26, 0.40]), and (b) the association between the "other posttraumatic stress measure" correlate and the HTQ-16, with "mental health setting" having the lowest coefficient (r = 0.28, p < 0.05, 95% CI [0.07, 0.47]) and "medical setting" the highest (r = 0.86, p < 0.001, 95% CI [0.67, 0.94]). The "type of trauma" variable was also a moderator for both of these convergent VG associations. Specifically, "intentional traumas" had the highest coefficient for the association between "cumulative trauma" and the HTQ-30 (r = 0.28, p < 0.001, 95% CI [0.20, 0.35], with a lower coefficient found for "unintentional traumas" (r = 0.04, p = 0.717, 95% CI [-0.17, 0.25]). However, for the association between "other posttraumatic stress measures" and the HTQ-16, the coefficient for "intentional traumas" was significantly lower (r = 0.58, p < 0.001, 95% CI [0.43, 0.70]) than the coefficient for "unintentional traumas" (r = 0.78, p < 0.001, 95% CI [0.65, 0.86]). Finally, for the association between the "cumulative trauma" correlate category and the HTQ-30, the coefficient for "original HTQ sample" was higher (r = 0.43, p < 0.001, 95% CI [0.26, 0.57]) than the coefficient for the "not original sample" (r = 0.21, p < 0.001, 95% CI [0.12, 0.29]).

Discussion

The goals of this study were to summarize the available data regarding the convergent and discriminant validity of the HTQ symptom scales. Of the 116 studies that provided validity estimates, 78 of them were included in the analyses. These studies represented 79 independent samples. The samples had very diverse compositions, with many languages, cultures, and countries of origin represented in the data. Although most authors recognized the need to adapt the HTQ, the manner in which they proceeded to make changes to the original did not generally

follow the adaptation recommendations set by the developers of this measure (i.e., culturally adapting the instrument, using blind-back consensus approach to translation, pre-testing the adapted version). These observations are similar to the studies included in the RG of Study 1 and reflect the frequent findings that (a) the process of translating and cross-culturally validating instruments is usually not a priority in clinical research protocols (Sperber, 2004) and (b) simple forward or back translations without a consensus approach are common in clinical cross-cultural research (Maneesriwongul & Dixon, 2004).

Overall, the results of the meta-analyses suggest that there is some evidence supporting the convergent validity of the HTQ-16. Unfortunately, the same cannot be said for the convergent validity of the HTQ-30. Because of the limited number of convergent correlates found in the research literature for this scale, there was only one category (i.e., cumulative trauma) that had sufficient data to compute effect size statistics. However, even for the HTQ-16, the results of the publication bias analysis indicated that substantially more research is required before one can consider that evidence for this measure's convergent validity has been established.

According to Hemphill's (2003) *r*-value effect size benchmarks for psychology, all the convergent validity correlates for the HTQ-16 had large effect sizes (i.e., $r > .30$). The strongest association was with cultural expressions of trauma and the weakest was with cumulative trauma. Although there were only three studies included in the cultural expressions of trauma analyses, the relatively high mean correlation indicates that the cultural symptoms endorsed by the samples represented in these studies (i.e., Quechuan/Peruvian, Cambodian, and West Papuan survivors of civil war) are closely related to the DSM-IV (APA, 1994) PTSD symptoms presented in the HTQ-16. However, because of the low power of these results we cannot assume

that cultural expressions of trauma in other cultures are also closely related to PTSD as measured by this scale. On the other hand, the high mean correlations found with other posttraumatic stress measures and dissociation measures provide evidence of the convergent validity of the HTQ-16 across various cultures. Similar to other PTSD measures (e.g., Dermichyan, Goenjian, & Khachadourian, 2015; Ruggiero, Del Ben, Scotti, & Rabalais, 2003), "cumulative trauma" had low correlations for both HTQ symptom scales. Because of variability in the conditional risk of PTSD across cultures, measures of traumatic events may not provide an optimal convergent validity correlate. In addition, it appears that trauma subtypes, rather than cumulative trauma, may better predict mental health outcomes (Arnetz et al., 2014).

With regards to the discriminant validity properties of the HTQ symptom scales, the results of the VG analyses are not as encouraging. More specifically, the evidence of discriminant validity of the HTQ-30 was rather weak. The associations between these related constructs and the HTQ-30 were very strong, especially with depressive symptoms with a mean correlation of .78. However, these results are not entirely surprising considering that several items of the HTQ-30 are characteristic of symptoms of depression (e.g. "hopelessness", "feeling guilty for having survived"). These findings suggest that the HTQ-30 may be a more general measure of psychological distress rather than a specific measure of PTSD symptoms.

In terms of the HTQ-16, all discriminant correlates of this scale had large effect sizes, according to Hemphill's (2003) benchmarks. Although it is expected that the HTQ-16 would have a certain degree of association with measures of depression and anxiety due to high rates of comorbidity and symptom overlap, the associations between these constructs and the HTQ-16 were quite strong (i.e., mean $r > .60$). However, these results appear to be comparable to other commonly used DSM-IV correspondent measures of PTSD, such as the Posttraumatic

Diagnostic Scale (PDS; Foa, Cashman, Jaycox, & Perry, 1997) and the PTSD Checklist-Civilian version (PCL-C; Weathers, Litz, Herman, Huska, & Keane, 1993). These instruments have shown variable performances in terms of discriminant validity in separate single studies, with some showing weaker associations than the HTQ-16 with constructs such as depression and anxiety (e.g., Adkins, Weathers, McDevitt-Murphy, & Daniels, 2008; Kornblith et al., 2003), and others showing very strong associations, especially with depression symptoms (i.e. $r > .70$) (e.g., Carter-Visscher et al., 2010; Foa et al., 1997). However, until meta-analytic investigations are also conducted with these instruments, it is difficult to assess whether the discriminant validity properties of the HTQ-16 are indeed statistically comparable.

Nevertheless, questions about the discriminant validity of PTSD measures have been raised for many instruments, as PTSD measures have been found to be highly correlated with measures of anxiety and, especially, depression symptoms (e.g., Foa et al., 1997). This pattern appears to be reflective of Miller, Fogler, Wolf, Kaloupek and Keane's (2008) model of psychiatric comorbidity that posits that PTSD shares more in common with depressive symptoms such as melancholy, apathy and rumination than it does with anxiety and avoidance. Consistent with this model, relatively weaker associations have been found between PTSD and alcohol abuse and psychopathy, than anxiety and depressive symptoms (e.g., Bovin et al., 2016), suggesting that these types of constructs may be more appropriate for examining the discriminant validity of a PTSD measure. Unfortunately, we were unable to examine this possibility, as there were so few reports in the literature of correlations between these specific constructs and the HTQ symptom scales.

In order to assess the overall construct validity of the HTQ-symptoms scales, it is important to consider the evidence of both the convergent and discriminant validity together.

Unfortunately, there was an insufficient amount of convergent validity correlates for the HTQ-30 to fully evaluate its construct validity. Although there are no specific benchmarks to evaluate and compare convergent and discriminant validity coefficients in meta-analysis, it is expected that similar constructs would have significantly higher correlations than would theoretically less related constructs (Campbell & Fiske, 1959). This is not the case for some correlates of the HTQ-16. For instance, the mean estimates for this scale with depressive symptom measures and other posttraumatic stress measures were almost identical, with considerable overlap among the confidence intervals for these two associations. The lowest mean estimates for the discriminant validity of the HTQ symptom scales were with somatization measures, but even here the values were relatively high. This strong association appears to be consistent with results from other research not involving the HTQ that have found high associations between PTSD and somatization (e.g., Beckham et al., 1998; North, Kawasaki, Spitznagel, & Hong, 2004). However, because other measures of DSM-IV PTSD have shown to have significantly higher correlations with other measures of posttraumatic symptoms (e.g. Ruggiero et al., 2003) than with measures of depression and anxiety symptoms, this calls into question the overall construct validity of the HTQ-16. Despite these findings, further data is needed before drawing definite conclusions, as the VG results in this study are not robust to publication bias.

Moderators

The statistically significant Q and high I^2 tests found with some of the mean validity estimates suggest that the variability among them is heterogeneous and is not due to sampling error alone. The high degree of variance that is caused by true heterogeneity of the mean estimates indicated that the examination of sample and methodological characteristics could help explain this variability.

Overall, the moderator analyses indicated that some study and sample characteristics had a moderating or predictive effect on both the HTQ symptom scales. In summary, cultural group, recruitment site, type of trauma, and whether the sample was a sample for which the HTQ was originally validated affected some of the convergent validity estimates (i.e. cumulative trauma and other posttraumatic stress measures). In terms of discriminant validity, age, gender, and cultural group played a role in the heterogeneity of the estimates for the somatization measures correlate category. Contrary to the hypotheses, neither having a sample comprised of forcibly displaced individuals nor the participants' country of origin's cultural orientation had a moderating effect on the convergent and discriminant validity of either scale.

Specifically regarding the HTQ-16, there were moderating effects when the scale was correlated with other posttraumatic stress measures and measures of somatization. More precisely, the convergent validity of the scale when correlated with other posttraumatic stress measures was higher with samples recruited in medical settings and samples that had experienced unintentional traumas. As for the discriminant validity of this scale when it is compared to somatization measures, lower estimates were obtained with older participants, samples with European backgrounds, and samples mostly comprised of men. These moderating effects appear to be consistent with other research that has shown that age, gender, and culture play a role in the development of PTSD (e.g., Ditlevsen & Elklit, 2010; Kessler et al., 1995; Kessler et al., 1999; Norris 1992; Pietrzak, Goldstein, Southwick, & Grant, 2011) and somatization (e.g., Kirmayer & Young, 1998; Leiknes, Finset, Moum, & Sandanger, 2007). Lower estimates were also found with samples of European background, meaning that the discriminant validity was weaker in non-European samples. Because of the high risk of publication bias and the low power of these moderating effects due to relatively small numbers

of studies included in the analyses, these results should be considered as preliminary; they do not indicate that there are significant limitations with the HTQ-16. However, further research is needed to make inferences regarding the convergent validity of the HTQ-16, especially in samples that have been recruited in community or mental health treatment settings, and with individuals who have experienced intentional traumas. Similarly, more research is needed to assess the discriminant validity of the HTQ-16 in younger samples, samples comprised mostly of women and with individuals of non-European backgrounds that had weaker results in this study.

As for the HTQ-30, moderator analyses could only be conducted with the cumulative trauma correlate and significant effects were found for this variable. The convergent validity estimates were higher in samples comprised of individuals of Asian backgrounds, in samples that were recruited in community settings, with individuals who had experienced intentional trauma, and with samples similar to those on which the HTQ was originally validated. These results are not surprising due to the fact that the 14 cultural items of the HTQ-30 were based on cultural expressions of trauma in Asian samples (specifically Indochinese populations) who had experienced intentional traumas such as war. These results further justify the importance of adapting the cultural items for the each specific culture. On the other hand, these estimates were high in samples recruited in community settings, whereas the original validation samples were recruited in mental health treatment settings. Once again, because of the risk of publication bias and low power of the moderator analyses, these findings should be interpreted with caution. However, researchers should note that it may be possible that the convergent validity of the HTQ-30 is compromised in non-Asian samples, samples that are recruited in clinical settings, samples that have experienced unintentional traumas, and samples for which the HTQ was not originally validated (i.e., non-Indochinese refugee populations).

Limitations of the Study

Some of the limitations of this study stem from the cited reference search strategy, that may not have identified all of the studies and dissertations that have used the HTQ symptom scales. In addition, the coding process was challenging at times because the information reported in the studies was often vague and difficult to locate. More specifically for the VG, problems of inconsistent reporting and lack of data were evident when extracting validity coefficients from identified studies. For instance, correlation values and/or matrices were not always available, and supplementary analyses were sometimes required to calculate the bivariate associations. In many cases, authors reported conducting bivariate analyses but did not provide sufficient information to include these findings in the meta-analyses.

Another limitation of this study comes from the grouping of estimates into larger correlate categories. Although this procedure helped increase statistical power, the correlates included in these broad categories were somewhat dissimilar because of the wide variety of instruments that were used to measure them. For instance, there were several depression symptom instruments that were used to measure the correlates included in the broad category of "Depression symptoms measures," such as the Indochinese versions of the HSCL-25 (Mollica et al., 1987) and the BDI-II (Beck et al., 1996). Thus, the correlate categories did not represent completely homogeneous constructs, which may have affected the results.

In addition, because the VG analyses included correlations, it is important to note that it was the ability of the HTQ symptom scales to assess posttraumatic symptom severity and not PTSD caseness that was evaluated. Future users of the HTQ in research settings are thus encouraged to evaluate the ability of the scale to diagnose PTSD by conducting receiver

operating characteristic (ROC) curve analyses, for instance, to establish appropriate cut-off scores for their specific sample.

Implications and Recommendations

In conclusion, several important implications and recommendations stem from this study. First, consistent with established publication standards, authors are strongly encouraged to report relevant and sufficient information in the methods and results sections of their study. As the accuracy of meta-analyses is dependent on the primary studies on which they are based, missing information and data in primary studies ultimately has a negative impact on the quality of meta-analyses.

Furthermore, the results obtained from the VG analyses focused on discriminant validity raise doubts about the overall construct validity of HTQ symptom scales and the construct validity of PTSD as conceptualized in the DSM. The main concern was the high degree of similarity between the convergent and discriminant validity estimates across a wide range of disparate variables. Although some studies support the hypothesis that removing symptoms of other constructs theorized to overlap with PTSD, such as mood and other anxiety disorders, increases the validity of the PTSD construct (e.g., Elhai et al., 2015; Spitzer, First, & Wakefield, 2007), others have found that removing these symptoms does not reduce comorbidity with these other disorders (e.g., Grubaugh, Long, Elhai, Frueh, & Magruder, 2010).

Given that PTSD is highly comorbid and has several symptoms that overlap with depression and anxiety, convergent and discriminant measures of PTSD should perhaps be mapped on a continuum, rather than be considered as distinct categories. Westen and Rosenthal (2003) proposed methods of quantifying construct validity based on contrast analysis that provide effect sizes that indicate the degree to which an observed pattern of convergent-

discriminant correlations is similar to a theoretically based predicted pattern of correlations. These pattern-mapping approaches to the evaluation of construct validity are particularly relevant in the case of PTSD, where patterns of relationships with other constructs have been identified in the literature (e.g., Bovin et al., 2015; Miller et al., 2008). Aggregating the findings of these studies could help establish theoretically based patterns of relationships between PTSD and other constructs and serve as benchmarks to interpret the magnitude of observed correlations.

Because the theoretical conceptualization of PTSD and the understanding of its relation to other constructs are in continuous development, it is difficult to evaluate whether poor psychometric properties of an instrument are truly a function of the scale, or rather the PTSD construct. Nonetheless, the aggregate findings of the VG analyses, although preliminary, indicate that the convergent validity properties of the HTQ-16 are supported to some extent, and that this scale may perform questionably in regards to discriminant validity. The present results, combined with the findings of other studies that have evaluated the validity of the HTQ symptom scales (e.g., Rasmussen et al., 2015), raise some doubts upon the overall construct validity of this scale. In addition, the moderator analyses suggest that caution should be taken in designing future studies that include variables that had significant moderating effects (e.g., convergent validity estimates were higher in samples that experienced unintentional traumas and samples that were recruited in medical treatment settings). Furthermore, there is limited support for either the convergent or discriminant validity of the HTQ-30. Because of the mixed results for the HTQ-16 and the fact that the HTQ symptom scales are commonly considered as strong cross-cultural measures of trauma symptoms in the literature, considerably more work is needed to establish the validity of these scales.

References

Adkins, J. W., Weathers, F. W., McDevitt-Murphy, M., & Daniels, J. B. (2008). Psychometric properties of seven self-report measures of posttraumatic stress disorder in college students with mixed civilian trauma exposure. *Journal of Anxiety Disorders*, *22*, 1393-1402. doi: 10.1016/j.janxdis.2008.02.002

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. revised). Washington, DC: Author.

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Arnetz, B. B., Broadbridge, C. L., Jamil, H., Lumley, M. A., Pole, N., Barkho, E., … Arnetz, J. E. (2014). Specific trauma subtypes improve the predictive validity of the Harvard Trauma Questionnaire in Iraqi refugees. *Journal of Immigrant and Minority Health*, *16*, 1055-1061. doi: 10.1007/s10903-014-9995-9

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Beckham, J. C., Moore, S. D., Feldman, M. E., Hertzberg, M. A., Kirby, A. C., & Fairbank, J. A. (1998). Health status, somatization, and severity of posttraumatic stress disorder in

Vietnam combat veterans with posttraumatic stress disorder. *American Journal of Psychiatry, 155*, 1565-1569. doi: 10.1176/ajp.155.11.1565

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive meta-analysis version 2*. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, UK: John Wiley & Sons, Ltd.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97-111. doi: 10.1002/jrsm.12

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition (PCL-5) in veterans. *Psychological Assessment*, 28, 1379-1391. doi: 10.1037/pas0000254

Briere, J., & Runtz, M. (1989). The Trauma Symptom Checklist: Early data on a new scale. *Journal of Interpersonal Violence, 4*(2), 151-163.

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: The 1996 Detroit Area Survey of Trauma. *Archives of General Psychiatry, 55*, 626– 631. doi: 10.1001/archpsyc.55.7.626

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletn, 56*, 81-105. doi: 0.1037/h0046016

Carter-Visscher, R., Polusny, M. A., Murdoch, M., Thuras, P., Erbes, C. R., & Kehle, S. M. (2010). Predeployment gender differences in stressors and mental health among U.S.

National Guard troops poised for Operation Iraqi Freedom deployment. *Journal of Traumatic Stress, 23*, 78-85. doi: 10.1002/jts.20481

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101–129. doi: 10.2307/3001666

Darzi, C. M., & Hunsley, J. (2017). *The Harvard Trauma Questionnaire: A reliability generalization study.* Manuscript in preparation.

Dermichyan, A., Goenjian, A. K., & Khachadourian, V. (2015). Factor structure and psychometric properties of the Posttraumatic Stress Disorder (PTSD) Checklist and DSM-5 PTSD symptom set in a long-term post-earthquake cohort in Armenia. *Assessment, 22,* 594-606. doi: 10.1177/1073191114555523

Dieckmann, N., Malle, B., & Bodner, T. (2009). An empirical assessment of meta- analytic practice. *Review of General Psychology, 13,* 101-115. doi:10.1037/a0015107

Ditlevsen, D. N., & Elklit, A. (2010). The combined effect of gender and age on posttraumatic stress disorder: Do men and women show differences in the lifespan distribution of the disorder? *Annals of General Psychiatry, 9*. doi: 10.1186/1744-859X-9-32

Dye, D. A., Reck, M., & McDaniel, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment, 1*(3), 153-157.

Elhai, J. D., Contractor, A. A., Tamburrino, M., Fine, T. H., Cohen, G., Shirley, E., …, Galea, S. (2015). Structural relations between DSM-5 PTSD and major depression symptoms in military soldiers. *Journal of Affective Disorders, 175*, 373-378. doi: 10.1016/j.jad.2015.01.034

Fazel, M., Wheeler, J., & Danesh, J. (2005). Prevalence of serious mental disorder in 7000 refugees resettled in western countries: A systematic review. *Lancet, 365,* 1309-1314.

doi: 10.1016/S0140-6736(05)61027-6

Foa, E., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of PTSD: The Posttraumatic Diagnostic Scale. *Psychological Assessment, 9,* 445-451. doi: 10.1037/1040-3590.9.4.445

Gagnon, A. J., Tuck, J., & Barkun, L. (2004). A systematic review of questionnaires measuring the health of resettling refugee women. *Health Care for Women International, 25,* 111-149. doi: 10.1080/07399330490267503

Goldberg, D., & Williams, P. (1988). *A user's guide to the General Health Questionnaire.* Windsor, UK: NFER-Nelson.

Grubaugh, A. L., Long, M. E., Elhai, J. D., Frueh, B. C., & Magruder, K. M. (2010). An examination of the construct validity of posttraumatic stress disorder with veterans using a revised criterion set. *Behavior, Research and Therapy. 48*, 909–914. doi: 10.1016/j.brat.2010.05.019

Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2011). *Scientific foundations of clinical assessment.* New York: Routledge.

Hemphill, J. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58,* 78-79. doi: 10.1037/0003-066X.58.1.78

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539-1558. doi: 10.1002/sim.1186

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the Mind,* (3rd ed.). New York: McGraw-Hill.

Hollifield, M., Warner, T. D., Lian, N., Krakow, B., Jenkins, J. H., Kesler, J., Stevenson, J., & Westermeyer, J. (2002). Measuring trauma and health status in refugees: A critical

review. *Journal of the American Medical Association, 288*, 611–621. doi: 10.1001/jama.288.5.611

Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41*, 209-218. doi: 10.1097/00006842-197905000-00004

Hunsley, J., & Mash, E. J. (2008). Developing criteria for evidence-based assessment: An introduction to assessments that work. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 3-14). New York: Oxford University Press.

Jones, E., Vermaas, R. H., McCartney, H., Beech, C., Palmer, I., Hyams, K., & Wessely, S. (2003). Flashbacks and posttraumatic stress disorder: The genesis of a 20th-century diagnosis. *British Journal of Psychiatry, 182,* 158–163. doi: 10.1192/bjp.182.2.158

Keane, T. M., Silberbogen, A. K., & Weierich, M. R. (2008). Posttraumatic stress disorder. In J. Hunsley & E.J. Mash (Eds.), *A guide to assessments that work* (pp. 293-315). New York: Oxford University Press.

Kelly, P. A., O'Malley, K. J., Kallen, M. A., & Ford, M. E. (2005). Integrating validity theory with use of measurement instruments in clinical settings. *Health Services Research, 40*, 1605-1619. doi: 10.1111/j.1475-6773.2005.00445.x

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry, 52*, 1048-1060. doi: 10.1002/1099-1298(200011/12)

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., Nelson, C. B., & Breslau, N. (1999). Epidemiological risk factors for trauma and PTSD. In R. Yehuda (Ed.), *Risk factors for posttraumatic stress disorder* (pp. 23-60). Washington, DC: American Psychiatric Press.

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments in research. *American Journal of Health-System Pharmacists, 65*, 2276-2284. doi: 10.2146/ajhp070364

Kinicki, A. J., McKee-Ryan, F. M., Schriescheim, C. A., & Carson, K. P. (2002). Assessing the construct validity of the Job Descriptive Index: A review and meta-analysis. *Journal of Applied Psychology, 87*, 14-32. doi: 10.1037//0021-9010.87.1.14

Kirmayer, L. J., & Young, A. (1998). Culture and somatization: Clinical, epidemiological, and ethnographic perspectives. *Psychosomatic Medicine, 60*(4), 420-430.

Kleijn, W. C., Hovens, J. E. & Rodenburg, J. J. (2001). Posttraumatic stress symptoms in refugees: Assessments with the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist-25 in different languages. *Psychological Reports, 88*, 527-532. doi: 10.2466/PR0.88.2.527-532

Kornblith, A. B., Herndon, J. E., Weiss, R. B., Zhang, C., Zuckerman, E. L., Rosenburg, S.,…Holland, J. C. (2003). Long-term adjustment of survivors of early-stage breast carcinoma, 20 years after adjuvant chemotherapy. *Cancer, 98*, 679-689. doi: 10.1002/cncr.11531

Lange, A., Rietdijk, D., Hudcovicova, M., van de Ven, J. P., Schrieken, B., & Emmelkamp, P. M. (2003). Interapy: A controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. *Journal of Consulting and Clinical Psychology, 71*, 901–909. doi: 10.1037/0022-006X.71.5.901

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal, 333*, 597-600. doi: 10.1136/bmj.333.7568.597

Leiknes, K. A., Finset, A., Moum, T., & Sandanger, I. (2007). Course and predictors of medically unexplained pain symptoms in the general population. *Journal of Psychosomatic Research, 62*, 119-128. doi: 10.1016/j.jpsychores.2006.08.009

Maneesriwongul, W. & Dixon, J. K. (2004) Instrument translation process: A methods review. *Journal of Advanced Nursing, 48*, 175– 186. doi: 10.1111/j.1365-2648.2004.03185.x

Matthieu, M., & Ivanoff, A. (2006). Treatment of human-caused trauma: Attrition in the adult outcomes research. *Journal of Interpersonal Violence, 2*, 1654–1664. doi: 10.1177/0886260506294243

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' response and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749. doi: 10.1037/0003-066X.50.9.741

Michalopoulos, L. M., Unick, G. J., Haroz, E. E., Bass, J., Murray, L. K., & Bolton, P. A. (2015). Exploring the fit of western PTSD models across three non-western low- and middle-income countries. *Traumatology, 21*, 55-63. doi: 10.1037/trm0000020

Miller, M. W., Fogler, J. M., Wolf, E. J., Kaloupek, D. G., & Keane, T. M. (2008). The internalizing and externalizing structure of psychiatric comorbidity in combat veterans. J*ournal of Traumatic Stress, 21*, 58–65. doi: 10.1002/jts.20303

Mollica, R. F., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., & Lavelle, J. (1992). The Harvard Trauma Questionnaire: Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *Journal of Nervous and Mental Disease, 180*, 111-116. doi: 10.1097/00005053-199202000-00008

Mollica, R. F., Wyshak, G., de Marneffe, D., & Lavelle, J. (1987). Indochinese versions of the Hopkins Symptom Checklist-25: A screening instrument for the psychiatric care of

refugees. *American Journal of Psychiatry, 144,* 497-500. doi: 10.1176/ajp.144.4.497

Nakeyar, C., & Frewen, P. A. (2016). Evidence-based care for Iraqi, Kurdish, and Syrian asylum seekers and refugees of the Syrian civil war: A systematic review. *Canadian Psychology, 57,* 233-245. doi: 10.1037/cap0000067

North, C. S., Kawasaki, A., Spitznagel, E. L., & Hong, B. A. (2004). The course of PTSD, major depression, substance abuse, and somatization after a natural disaster. *Journal of Nervous and Mental Disease, 192,* 823-829. doi: 10.1097/01.nmd.0000146911.52616.22

Norris, F. H. (1992). Epidemiology of trauma: Frequency and impact of different potentially traumatic events on different demographic groups. *Journal of Consulting and Clinical Psychology, 60,* 409-418. doi: 10.1037/0022-006X.60.3.409

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality, 17,* S19-S38. doi: 10.1002/per.487

Orwin, R. G. (1983). A fail-safe *N* for effect sizes in meta-analysis. *Journal of Educational Statistics, 8,* 157-159. doi: 10.3102/10769986008002157

Perkonigg, A., Kessler, R. C., Storz, S., & Wittchen, H.-U. (2000). Traumatic events and post-traumatic stress disorder in the community: Prevalence, risk factors and comorbidity. *Acta Psychiatrica Scandinavica, 101,* 46-59. doi: 10.1034/j.1600-0447.2000.101001046.x

Pietrzak, R. H., Goldstein, R. B., Southwick, S. M., & Grant, B. F. (2011). Medical comorbidity of full and partial posttraumatic stress disorder in US adults: Results from Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychosomatic Medicine, 73,* 697-707. doi: 10.1097/PSY.0b013e3182303775

Porter, M., & Haslam, N. (2005). Pre-displacement and post-displacement factors associated with mental health of refugees and internally displaced persons: A meta-analysis. *Journal of the American Medical Association, 294*, 602-612. doi: 10.1001/jama.294.5.602.

Rasmussen, A., Verkuilen, J., Ho, E., & Fan, Y. (2015). Posttraumatic stress disorder among refugees: Measurement invariance of Harvard Trauma Questionnaire scores across global regions and response patterns. *Psychological Assessment, 27*, 1160-1170. doi: 10.1037/pas0000115

Ruggiero, K. J., Del Ben, K., Scotti, J. R., & Rabalais, A. E. (2003). Psychometric properties of the PTSD Checklist-Civilian version. *Journal of Traumatic Stress, 16,* 495-502. doi: 10.1023/A:1025714729117

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529-540. doi: 10.1037/0021-9010.62.5.529

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274. doi: 10.1.1.172.1733

Shoeb, M., Weinstein, H., & Mollica, R. (2007). The Harvard Trauma Questionnaire: Adapting a cross-cultural instrument for measuring torture, trauma and posttraumatic stress disorder in Iraqi refugees. *The International Journal of Social Psychiatry, 53,* 447-463. doi: 10.1177/0020764007078362

Sperber, A. D. (2004) Translation and validation of study instruments for cross-cultural research. *Gastroenterology, 126,* S124–S128. doi: 10.1053/j.gastro.2003.10.016

Spitzer, R. L., First, M. B., & Wakefield, J. C. (2007). Saving PTSD from itself in DSM-V. *Journal of Anxiety Disorders, 21,* 233-241. doi: 10.1016/j.janxdis.2006.09.006

Steel, Z., Chey, T., Silove, D., Marmane, C., Bryant, R. A., & van Ommeren, M. (2009). Association of torture and other potentially traumatic events with mental health outcomes among populations exposed to mass conflict and displacement: A systematic review and meta-analysis. *Journal of the American Medical Association, 302,* 537-549. doi: 10.1001/jama.2009.1132

Van der Velden, P. G., Grievink, L., Kleber, R. J., Drogendijk, A. N., Roskam, A. J., Marcelissen, F. G., … Gersons, B. P. (2006). Post-disaster mental health problems and the utilization of mental health services: A four-year longitudinal comparative study. *Administration and Policy in Mental Health, 33,* 279–288. doi: 10.1007/s10488-005-0027-x

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD checklist (PCL): Reliability, validity, and diagnostic utility.* Poster presented at the 9th annual meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology, 84,* 608-618. doi: 10.1037/0022-3514.84.3.608

Wood, J. (2008). Methodology for dealing with duplicate study effects in meta-analysis. *Organizational Research Methods, 11,* 79-95. doi: 10.1177/1094428106296638

Figure 1

Literature Search Flow Chart

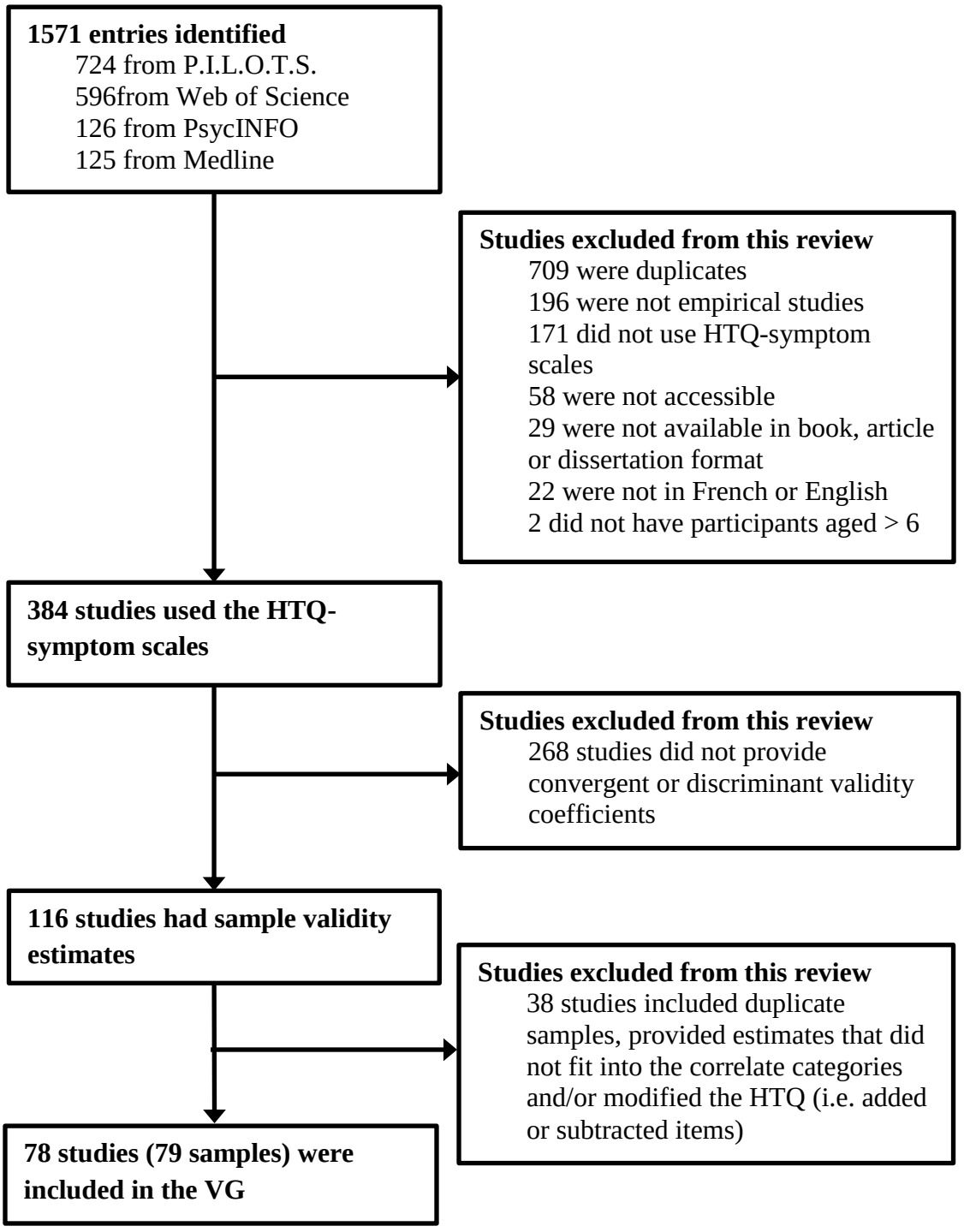

Table 1

Descriptive Statistics for the Convergent Validity of the HTQ Symptom Scales

Validity estimate	k	Mean *r*	*95% confidence interval* Lower	Upper	Min.	Max.	*Q*	I^2	*Orwin's N*
Cumulative trauma									
HTQ-16	43	.32	.26	.38	-.28	.90	546.53*	92.32	17
HTQ-30	12	.26	.17	.34	.04	.53	48.16*	77.16	1
Posttraumatic stress measures									
HTQ-16	11	.66	.52	.76	.24	.86	154.92*	93.55	41
Cultural expressions of trauma									
HTQ-16	3	.75	.59	.86	.55	.83	29.38*	93.19	15
Dissociation Measures									
HTQ-16	3	.53	.32	.70	.18	.69	33.46*	94.03	8

Note. Min. = Minimum correlation reported in the literature; Max. = Maximum correlation reported in the literature; *k* = number of samples included in the analyses; N/A = Insufficient number of samples to compute Orwin's *N*
*p < .001, **p < .01

Table 2

Descriptive Statistics for the Discriminant Validity of the HTQ Symptom Scales

Validity estimate	k	Mean *r*	*95% confidence interval* Lower	Upper	Min.	Max.	*Q*	I^2	*Orwin's N*
Depressive and anxiety symptoms measures									
HTQ-16	3	.80	.77	.83	.79	.86	1.73	0	15
HTQ-30	2	.79	.61	.89	.71	.85	12.84*	92.21	N/A
Anxiety symptoms measures									
HTQ-16	12	.62	.56	.68	.44	.78	61.47*	82.10	31
HTQ-30	2	.66	.16	.89	.43	.80	18.49*	94.60	N/A
Depression symptoms measures									
HTQ-16	21	.64	.57	.69	.16	.84	199.78*	90.00	58
HTQ-30	5	.78	.72	.82	.64	.82	9.02	55.67	23
Somatization measures									
HTQ-16	6	.43	.28	.56	.17	.62	47.11*	89.37	11
HTQ-30	2	.45	-.31	.86	.08	.72	37.24*	97.32	N/A

Note. Min. = Minimum correlation reported in the literature; Max. = Maximum correlation reported in the literature; *k* = number of samples included in the analyses; N/A = Insufficient number of samples to compute Orwin's *N*
*p < .001

Table 3

Summary of Moderating Effects for the Construct Validity of the HTQ Symptom Scales

	Cumulative trauma		PTSD measures	Anxiety symptoms measures	Depressive symptoms measures	Somatization measures
Moderator	HTQ-16 *N* = 43	HTQ-30 *N* = 12	HTQ-16 *N* = 11	HTQ-16 *N* = 12	HTQ-16 *N* = 21	HTQ-16 *N* = 6
Age	No (34*)	No (10)	No (8)	No (10)	No (18)	Yes (5)
Gender	No (40)	No (12)	No (9)	No (12)	No (20)	Yes (6)
Cultural group	No (34)	Yes (9)	No (10)	No (11)	No (19)	Yes (5)
Cultural orientation	No (19)	No (5)	No (5)	No (4)	No (9)	N/A
Forcibly displaced sample	No (43)	No (12)	No (11)	No (11)	No (20)	No (5)
Recruitment site	No (41)	Yes (12)	Yes (10)	No (11)	No (20)	No (6)
Trauma type	No (39)	Yes (12)	Yes (11)	N/A	No (19)	N/A
Original sample	No (43)	Yes (12)	Yes (12)	No (12)	No (21)	N/A

Note. *Total samples included in the moderator analyses are in parentheses; Yes = Significant moderating effect; No = No significant moderating effect; N/A = Insufficient information available to conduct moderator analyses

General Discussion

The Harvard Trauma Questionnaire (HTQ; Mollica et al., 1992) is a measure that was intended to assess traumatic events and related symptoms of trauma across cultures. The HTQ has been found to be commonly used in the literature (Gagnon et al., 2004; Hollifield et al., 2002), and has been shown to be useful to exclude PTSD non-cases in certain forcibly displaced populations (Nakeyar & Frewen, 2016). This instrument also has a reputation of being a culturally sensitive measure of PTSD symptoms (Keane et al., 2008). However, to date, there has been no thorough review that supports this reputation and evaluates the overall psychometric strength of the HTQ symptom scales as measures of PTSD across cultures. This dissertation was designed to fill this gap by: (a) gathering descriptive information regarding the use of the HTQ symptom scales in the empirical literature, (b) evaluating the reliability properties of this measure by calculating mean internal consistency estimates, (c) evaluating whether the HTQ symptom scales have produced valid scores across studies found in the literature by calculating mean construct validity estimates, and (d) examining sample and methodological characteristics that may influence the reliability and construct validity of scores produced by these scales. The following sections include a summary of both meta-analytic studies included in this dissertation, as well as the implications and recommendations stemming from their findings while taking into account potential limitations.

The Psychometric Properties of the HTQ

The literature search of studies that used the HTQ symptom scales in Study 1 shows no signs of a decline in the frequency of use of the scales. Researchers across the globe have used the HTQ to assess posttraumatic stress symptoms in refugee and nonrefugees, clinical and nonclinical populations, and across a wide range of cultural and ethnic backgrounds. Researchers

have also made attempts at adapting the HTQ for their specific sample. However, in the majority of studies, the reported adaptation procedures they used did not follow the translation and cultural adaptation guidelines provided by the original developers, nor did they follow commonly known adaptation recommendations (e.g., International Test Commission, 2005). Although the current study could not evaluate the efficacy of following these adaptation procedures, several studies offer support for the use of carefully implemented adaptation strategies in cross-cultural assessment to help minimize bias (e.g. Dolnicar & Grün, 2007, Harzing, 2006). Furthermore, those who used the HTQ-30 used the Indochinese version of the scale. This scale was based on the cultural expressions of trauma among Cambodian, Lao and Vietnamese refugees that may not be appropriate for individuals of other cultures. Although some researchers had translated the cultural items for use in their studies (e.g., Vloeberghs, van der Kwaak, Knipscheer, & ven den Muijsenbergh, 2012), the sense or meaning of the cultural items remained the same, even when participants did not have Cambodian, Lao, or Vietnamese backgrounds. However, because there was no opportunity to evaluate the 14 cultural items of the HTQ-30 due to insufficient data, it is difficult to determine if the inclusion of these items enhances the use of the HTQ across cultures.

To assess the reliability properties of the HTQ symptom scales, I conducted reliability generalization analyses on Cronbach's alpha coefficients in Study 1. These coefficients were retrieved from the published literature and unpublished dissertations, or were provided by authors after requesting this information. There was sufficient data to conduct RG analyses for the HTQ-16, HTQ-30, and the re-experiencing, avoidance and arousal subscales. Publication bias analyses revealed that all RG results were likely to be unaffected by this form of bias. Taking into account the results of the Spearman-Brown formula, the HTQ-16 and HTQ-30 had

comparable mean coefficients that would be described as "excellent" (i.e., ≥ .90) according to Hunsley and Mash's (2008) guidelines. Although both had excellent ratings, the addition of the cultural expressions of trauma items did not add value to the HTQ-16 in terms of increasing internal consistency. Regarding the subscales of the HTQ-16, the Spearman-Brown formula was used to compare the subscales with each other. The re-experiencing and arousal subscales had a rating of "good" (i.e., mean values in the.80-.89 range) and the avoidance/numbing subscale a rating of "adequate" (i.e., mean values in the .70-.79 range). In summary, the RG analyses revealed that the reliability properties of the HTQ-30, HTQ-16, and the re-experiencing and arousal subscales are strong. However, the reliability properties of the avoidance/numbing subscale are weaker. The lower level of reliability for this subscale may be reflective of an inadequate set of items chosen to represent avoidance/numbing symptoms, or even the poor factorial validity of the DSM-IV 3-factor model of PTSD (e.g. Yufik & Simms, 2010). Another potential cause for the suboptimal performance of the numbing subscale may be the inconsistent presence of avoidance and numbing symptoms across cultures (Marsella, Friedman, Gerrity, & Scurfield, 1996; Norris, Van Landingham, & Vu, 2009). This issue was highlighted throughout the revision process for the new DSM-5 criteria for PTSD (Hinton & Lewis-Fernández, 2010), during which it was decided to separate the avoidance and numbing symptoms into separate clusters. To date, there is no updated HTQ available that reflects the changes made to the PTSD construct in DSM-5. If these changes were made, perhaps the reliability of the subscales would improve. Therefore in terms of reliability properties, the two HTQ symptom scales as a whole appear to be strong, however some caution should be used when interpreting the subscales.

To assess the construct validity of the HTQ symptom scales in Study 2, I conducted validity generalization analyses on convergent and discriminant validity coefficients retrieved

from the published literature and unpublished dissertations. The coefficients were grouped according to broader correlate categories. The convergent validity correlate categories were: (a) cumulative trauma; (b) other posttraumatic stress measures; (c) cultural expressions of trauma; and (d) dissociation measures. The discriminant validity categories were: (a) depression and anxiety symptoms measures; (b) anxiety symptoms measures; (c) depression symptoms measures; and (d) somatization measures. Publication bias analyses revealed that the overall VG findings should be only considered as preliminary until additional data are available.

Based on Hemphill's (2003) *r*-value effect-size benchmarks, there was strong evidence for the convergent validity of the HTQ-16, as all correlates had large effect sizes (i.e., > .30). The strongest association was with cultural expressions of trauma and the weakest with cumulative trauma. The mean correlation between the HTQ-30 and cumulative trauma was considered a medium effect size (i.e., $r = .26$). There were insufficient data to compute VG estimates for the other convergent correlate categories for the HTQ-30. Similar to other PTSD measures (e.g., Dermichyan, Goenjian, & Khachadourian, 2015; Ruggiero, Del Ben, Scotti, & Rabalais, 2003), the cumulative trauma category demonstrated the lowest convergent validity for the HTQ-16 and only had a medium effect for the HTQ-30. Because of variability in the conditional risk of PTSD across cultures, traumatic events may not be an ideal convergent validity correlate. In addition, it appears that trauma subtypes can predict mental health outcomes better than cumulative trauma (Arnetz et al., 2014), which could be a reason for the weaker association with PTSD. Because of these reasons, future validation studies of PTSD measures should refrain from using this variable as a convergent validity correlate.

The results of Study 2 call into question the overall discriminant validity properties of the HTQ symptom scales. Based on Hemphill's (2003) guidelines, all discriminant correlates of both

scales had large effect sizes. Similar to other PTSD measures (e.g., Bovin et al., 2016; Dermichyan et al., 2015; Lauterbach, Vrana, King, & King, 1997), the associations between the HTQ symptom scales and depression and anxiety symptom measures were rather strong (i.e., > .60). Although large correlations were expected due to the high rates of comorbidity and symptom overlap with depression and anxiety, PTSD measures should have significantly higher correlations with similar constructs than with theoretically less related constructs. This was not the case for the HTQ-16 where some mean discriminant validity coefficients were comparable in size to convergent validity coefficients. Furthermore, although somatization measures had lower mean correlations with both HTQ symptom scales, they were still relatively high. This strong association is consistent with the results of other studies showing high associations between PTSD and somatization (e.g., Beckham et al., 1998; North, Kawasaki, Spitznagel, & Hong, 2004). The present results, combined with the findings of other studies that have evaluated the validity of the HTQ symptom scales (e.g., Rasmussen et al., 2015), indicate that the construct validity of both scales appears to be "adequate," based on Hunsley and Mash's (2008) criteria for psychometric ratings. In other words, there is some independently replicated evidence of construct validity for the scales.

In summary, the findings of the RG and VG analyses indicate that the HTQ-16 is likely to produce reliable scores across various types of samples. This scale also has some evidence of convergent validity, but performs rather poorly with regards to discriminant validity. In terms of the subscales of the HTQ-16, the reliability properties of the re-experiencing and arousal subscales were good, but the avoidance/numbing subscale was only considered adequate. With regard to the HTQ-30, it is important to note that the majority of studies that used this scale did not report culturally adapting the HTQ for their specific sample. These studies thus applied the

Indochinese version of the questionnaire to non-Indochinese groups. Although the cultural items that are based on clinical findings among Indochinese populations provided reliable scores across various populations, the ability to discriminate from measures of anxiety and depression symptoms was weak. There was insufficient information to evaluate the overall convergent validity, which prevented from fully assessing the construct validity of this scale.

The Role of Moderators

The majority of mean reliability and validity coefficients calculated from the RG and VG analyses had statistically significant *Q* and high I^2 tests. These results indicated that these estimates had heterogeneous variability, which was not due to sampling error alone, and that the examination of sample and methodological characteristics could perhaps explain this variability. Meta-regression and subgroup analyses were thus performed on reliability and validity coefficients that had a statistically significant *Q*-statistic and sufficient amount of available data. There were a total of eight variables that were selected as potential moderators, and seven of these variables played a moderating role with the reliability and/or validity estimates. These included age, gender, cultural group, cultural orientation, recruitment site, trauma type and whether the study sample was a sample for which the HTQ was originally validated.

Having the experience of being a forcibly displaced person did not have a significant moderating effect, suggesting that the HTQ symptom scales and subscales should perform rather equally in terms of psychometric properties in both forcibly displaced samples and not forcibly displaced samples. This is somewhat surprising given that the HTQ symptom scales were originally validated for forcibly displaced samples (i.e., refugees). However, there may be underlying factors that account for the lack of a moderating effect. For instance, studies have found that several elements may a role in the association between the experience of being

forcibly displaced and psychological distress, such as pre-migratory potentially traumatic events and post-migration stressors (e.g. Laban, Gernaat, Komproe, Van, & de Jong, 2005; Schwarz-Nielsen & Elklit, 2009), as well as the length of stay in the host country/area (Guajardo, Slewa-Younan, Smith, Eagar, & Stone, 2016). Because the forcibly displaced samples included in the current analyses varied greatly with regards to these factors, the heterogeneous characteristic of these samples may explain the lack of a moderating effect of this variable.

In terms of significant moderating effects, the mean age of samples played a role in the discriminant validity of the HTQ-16, specifically with the associations between this scale and somatization measures. Because higher associations were found with lower mean ages of study samples, it appears that the discriminant validity of the HTQ-16 may be compromised when it is compared to somatization measures in younger age groups. This moderating effect appears to support other findings that have shown significant associations between age and PTSD (e.g., Kessler et al., 1995; Norris 1992), as well as age and somatization (e.g., Leiknes, Finset, Moum, & Sandanger, 2007).

Gender also played a moderating role with some reliability and discriminant validity estimates. Specifically, study samples comprised of mainly female participants had higher reliability estimates on the re-experiencing subscale than did samples comprised mainly of male participants, as well as with the associations between the HTQ-16 and somatization measures. This is seemingly consistent with the findings of other studies that have shown gender differences in somatization (e.g., Leiknes et al., 2007) and PTSD (e.g., Ditlevsen & Elklit, 2010; Pietrzak et al., 2011), including re-experiencing symptoms specifically (Vassar, Knaup, Hale, & Hale, 2011). These results suggest that the discriminant validity of the HTQ-16 and the reliability of the re-experiencing subscale may be weaker in male samples.

Although it was expected that the cultural group would have a significant moderating effect on all of the dimensions of reliability and construct validity examined, this variable only had an impact on some aspects. Specifically, the cultural group had a predictive effect on the reliability of scores on the re-experiencing subscale, the convergent validity of the HTQ-30 (i.e., association with cumulative trauma) and the discriminant validity of the HTQ-16 (i.e., association with somatization measures). These findings support the existing evidence of cultural and ethnic variations in PTSD symptomatology (e.g., Kessler et al., 1999), exposure to traumatic events (e.g., Roberts, Gilman, Breslau, Breslau, & Koenen, 2011) and somatization (e.g., Kirmayer & Young, 1998).

As an extension to this, the cultural background of the samples' country in terms of the construct of collectivism versus individualism played a moderating role in the reliability of scores of the HTQ. Specifically, the reliability coefficient of the avoidance/numbing subscale for samples from individualist countries was significantly higher than the typical estimate for samples from collectivist countries. This may be reflective of lower rates of avoidance/numbing symptoms reported in collectivist cultures (e.g., McCall & Resick, 2003). These results also support Jayawickreme, Jayawickreme, and Foa's (2012) hypothesis that individuals from collectivist cultures may have less avoidant behaviours because of social roles and obligations that require them to consistently engage with others and their environment. This variable however, did not have a significant impact on the validity coefficients. Given the small amount of empirical evidence to date on this matter, it is clear that further research is needed to makes inferences regarding the role of cultural orientation on posttraumatic symptoms.

Similar to the cultural group variable, the type of trauma experienced (i.e., intentional versus unintentional) played a role in both the reliability and construct validity of the HTQ. The

reliability of scores produce by the re-experiencing subscale was stronger for “intentional traumas.” Furthermore, the experience of “intentional traumas” was also associated with stronger convergent validity properties of the HTQ-30. As intentional traumas are considered to be more severe than non-intentional traumas, these results may indicate that the internal consistency of re-experiencing symptoms and the convergent validity of the HTQ-30 are associated with trauma severity. On the other hand, conflicting results were obtained for the HTQ-16 that had stronger convergent validity properties with “unintentional traumas.” These results indicate that further research is needed regarding the role of trauma types and trauma severity to further test these interpretations.

The sites from which the samples were recruited had a significant effect on the construct validity of the HTQ symptom scales. The convergent validity of the HTQ-30 when associated with cumulative trauma was stronger for samples recruited in community settings. In addition, the convergent validity of the HTQ-16 when associated with other posttraumatic stress measures was stronger for samples recruited in medical treatment settings. Although the underlying factors contributing to these differences are unclear, these findings provide further support regarding the questionable generalizability of the scales due to variability in the sensitivity across recruitment settings (e.g., Mollica et al. 1992; Silove et al., 2007). Further studies are needed to clarify these findings.

Finally, being a sample for which the HTQ was originally validated had a positive effect on the association between cumulative trauma and the HTQ-30. These results provide some (although rather limited) support for the convergent validity of this scale in Indochinese refugee populations.

In summary, no moderators played a role in the overall reliability of scores of the HTQ-16 and HTQ-30. This finding suggests that these scales perform relatively consistently across various sample type and methodological procedures. The scales did differ, however, in terms of moderators for the construct validity. Because of the high risk of publication bias and the low power of these moderating effects due to relatively small numbers of studies included in the analyses, these results do not indicate that there are significant limitations with either HTQ symptom scale and should only be considered as preliminary findings.

Limitations of the Studies

There are several limitations to these studies. As highlighted in both studies, the cited reference search strategy used to identify published studies may not have identified all published studies and unpublished dissertations that have used the HTQ symptom scales. In addition, this strategy could not identify studies, other than dissertations, that have used this measure but are unpublished. This is important to bear in mind given that the publication bias analyses for the VG indicated that few studies would be needed to significantly lower the mean coefficients reported in the analyses. Thus, the findings of the VG in particular should be considered as preliminary until considerably more construct validity data is available on the HTQ scales.

For the studies identified with the search strategy, data from many were not included in the RG and VG analyses because reliability and validity estimates were either not reported in the primary studies or, as in the case for the VG, authors reported conducting bivariate analyses but did not provide sufficient information to include these findings in the meta-analyses. Although some authors provided a coefficient alpha after being contacted, there were still many missing reliability estimates that could have been included in the RG.

Furthermore, the coding process used to extract information for the analyses was challenging at times given that the information reported in the studies was often vague, challenging to interpret, and difficult to locate. A large amount of the information coded was retrieved in the methods sections where the adaptation and administration procedures of the HTQ would normally be described. However, there was often a lack of sufficient detail reported on these procedures. It is therefore possible that the information reported in these studies did not fully represent the actual methodological procedures, including cross-cultural adaptations that were used in the research. Indeed, the lack of details in the reporting of adaptation methods appears to be a common theme in the published literature (Maneesriwongul & Dixon, 2004). These problems continue despite reporting standards for research that been put in place, in part, to ensure the quality of research synthesis and meta-analysis (Appelbaum, Cooper, Maxwell, Stone, & Sher, 2008).

Another possible limitation lies in the merging of estimates of potential moderators and VG correlates into larger categories. Although this procedure helped increased statistical power, there may have been an important loss of information, and perhaps less accurate results. For instance, combining individual cultural groups into a broader category (e.g., combining Cambodian and Tibetan samples into the Asian category) considers these groups as homogenous when in fact there can be important differences across subgroups. This procedure thus reduces the generalizability of the findings. Moreover, the merging of VG estimates into broad correlate categories involved combining estimates provided by several different instruments that may not have been measuring the same underlying construct. For instance, there were several instruments, such as the Indochinese versions of the Hopkins Symptom Checklist-25 (Mollica et al., 1987) and the Beck Depression Inventory-II (Beck, Steer, & Brown, 1996) that were

included in the broad category of "depression symptoms measures." One way of ensuring that these instruments were indeed measuring the same construct was to evaluate the psychometric properties of each instrument. However, because of the numerous measures included, the psychometric evaluation of each of them was not feasible.

Furthermore, only 20% of the studies included in each of the RG and VG analyses were coded by a second coder. Although the interrater reliability was rated as "good" for both studies, a higher percentage of studies coded by the second coder may have increased the accuracy of the overall RG and VG results.

Finally, the decision to include samples consisting of children and adolescents by using a lower age limit of 7 years may not have been the most judicious. The variations in the symptom expression of PTSD due to developmental factors, especially in school-aged children (e.g. Cohen & the Work Group on Quality Issues, 1998), could have impacted the accuracy of the results. However, there was only one sample in the RG and two samples in the VG that included children as young as 9 years. Although approximately 23% of the RG samples and 20% of the VG samples included adolescents, PTSD symptoms during this developmental stage are more likely to be similar to symptoms experienced in adulthood (Cohen & the Work Group on Quality Issues, 1998). Because there were only a few studies that included younger children, the possible negative impact on the overall meta-analytic results is likely to be minimal.

Implications

In addition to the information regarding the psychometric properties of the HTQ symptom scales, several other important implications stem from the current studies. For instance, the findings of both studies add fuel to the debate regarding etic versus emic approaches for the assessment of PTSD. The findings revealed that an etic approach to assessment across cultures

can be justified given that the HTQ-16 produced reliable scores and had sound convergent validity properties across diverse samples. In addition, the variability in the reliability of scores among the subscales (i.e., re-experiencing, avoidance/numbing, arousal symptoms) indicates that the etic approach to the assessment of PTSD can help capture the cultural nuances in the expression of symptom sets when the same measure is used cross-culturally. On the other hand, the practice of incorporating an emic perspective to the etic approach is also supported, as the HTQ-30 had excellent reliability properties as well as some evidence of convergent validity (especially in samples similar to those on which the HTQ was originally validated). Although there were no data available to compare strictly emic approaches to assessment, the findings of the current studies suggest that there is no right or wrong answer to this debate, and that each approach can be justified to be used for specific sets of purposes. For instance, the etic approach may be more useful to evaluate cross-cultural differences in research settings, whereas an emic approach would be more appropriate to help better understand the expressions of trauma in a certain cultural group. A combination of emic and etic approaches to the assessment of PTSD would be particularly useful in clinical settings. The etic approach could help guide clinicians in establishing a diagnosis comparing symptom patterns to normative data, whereas the information provided by emic approaches could help tailor treatment interventions according to the specific needs of a client.

Furthermore, the results obtained from the VG analyses focused on discriminant validity raise doubts about the overall construct validity of HTQ symptom scales and the construct validity of PTSD as conceptualized in the DSM-IV (APA, 1994). The main concern was the high degree of similarity between the convergent and discriminant validity estimates across a wide range of disparate variables. Although some studies support the hypothesis that removing

symptoms of other constructs theorized to overlap with PTSD, such as mood and other anxiety disorders, increases the validity of the PTSD construct (e.g., Elhai et al., 2015; Spitzer, First, & Wakefield, 2007), others have found that removing these symptoms does not reduce comorbidity with these other disorders (e.g., Grubaugh, Long, Elhai, Frueh, & Magruder, 2010). In attempts to find a solution to this debate, some researchers have gone further and suggested that taking a dimensional approach to PTSD assessment is more appropriate than conceptualizing the disorder into a category (e.g., Antony & Rowa, 2005). For instance, cluster and factor analytic studies have found interesting interactions between PTSD and the internalizing and externalizing dimensions to psychopathology (Cox, Clara, & Enns, 2002; Miller, Fogler, Wolf, Kaloupek, & Keane, 2008). Clearly, more research is needed to resolve this key validity issue for both the construct of PTSD and measures of PTSD across diverse populations.

Although the HTQ symptom scales were based on DSM-IV PTSD criteria, the findings of the current studies also have some implications for the PTSD construct as conceptualized in DSM-5. This most recent version includes an even broader set of symptoms with an addition of three symptoms that were not included in DSM-IV and an increased emphasis on depressive mood and cognition (Friedman, 2013). Because the current findings question the DSM-IV PTSD construct due to a lack of ability to discriminate PTSD symptoms from symptoms of depression and other anxiety disorders, it is possible that an increase in dysphoria symptoms in the DSM-5 would render the new conceptualization of PTSD even more questionable. Recent studies that have used DSM-5 based measures suggest that the dysphoria symptoms of PTSD may be the underlying reason for the observed high level of comorbidity with major depressive disorder (Contractor et al., 2014; Elhai et al., 2015). However, these studies were conducted with North American samples and it is therefore not appropriate to generalize these findings to other

cultures. On the other hand, one study comparing the DSM-IV and DSM-5 conceptualization of PTSD found no significant differences in comorbidity rates with major depressive episode (O'Donnell et al., 2014) revealing no great impact of the addition of mood related items to PTSD in the DSM-5. This study would also need to be replicated across diverse cultural groups to evaluate the generalizability of these findings. Also, because of the limitations of Study 2 (i.e., the potential for publication bias, low power), these interpretations should be considered tentative. Nevertheless, the findings of the current studies further justify the need for a deeper investigation into the cross-cultural applicability of the PTSD construct and of PTSD measures.

Recommendations for Researchers

Several important recommendations for researchers stem from this study. First, consistent with established publication standards, authors are encouraged to report relevant and sufficient information in the methods and results sections of their research publications. The inconsistent and vague reporting of adaptation procedures, insufficient information to compute validity coefficients, and low rate in the reporting of reliability estimates are indications that more awareness is needed with regards to reporting guidelines for assessment instruments and assessment research. As the accuracy of meta-analyses is dependent on the primary studies on which they are based, missing information and data in primary studies ultimately has a negative impact on the quality of meta-analyses. Researchers are thus encouraged to follow established reporting guidelines such as the journal article reporting standards (JARS) when writing manuscripts. Cooper (2011) has provided a practical guide on understanding and implementing these standards. Although researchers may be aware of these standards, editorial policies can prevent them from following these standards by limiting the amount of information allowed in manuscripts. If there are space limitations, editors are encouraged to follow the example of some

journals, such as the journals of the American Psychological Association, that provide external links to websites where supplementary information can be viewed (American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008). In light of the challenges in the data retrieval of the current meta-analyses, researchers are particularly encouraged to provide the reliability estimates of their sample, detailed methodological procedures (including instrument adaptation procedures, if applicable), and correlation matrices. The reporting of this information is important not only for meta-analytic purposes, but also for helping readers to evaluate the research and to decide whether a measure is appropriate for their specific use.

The current studies also highlighted significant problems in the utilization practices of the HTQ symptom scales among researchers. Users of the HTQ (and any other assessment instrument, for that matter) must be vigilant of the potential sources of bias, especially in samples of different cultural backgrounds for which the original version of the measure was intended to be used, and carefully address bias by ensuring measurement equivalency. This includes evaluating whether an instrument can be applicable to a specific culture and if necessary, using appropriate translation and cultural adaptation strategies Although there is a lack of consensus regarding the specific adaptation procedures to be used (Epstein et al., 2015), all experts in cross-cultural measurement agree that the adaptation process goes beyond a mere translation. Common recommendations include using a consensus approach to translation, using focus groups and validating the adaptation (Epstein et al., 2015). Cross-cultural validation strategies can include multi-group confirmatory analyses and differential item functioning (Milfont & Fischer, 2010). Although the factorial validity of the HTQ was not evaluated in the current studies, the variable reliability properties of the HTQ-16 subscales, further support the

need for factor analyses in different cultural groups. Future users of the HTQ in research settings are also encouraged to evaluate the ability of the scale to diagnose PTSD by conducting receiver operating characteristic (ROC) curve analyses to establish appropriate cut-off scores for their specific sample. The high cost and lengthy process of adapting and validating instruments can be major obstacles in conducting research, but because the assessment of traumatized individuals can have major clinical and societal implications (e.g., access to treatment, obtaining disability compensation, obtaining refugee status), these procedures should not be ignored.

Recommendations for Clinicians

The aggregate nature of the findings in the current meta-analytic studies makes it challenging to provide precise clinical recommendations regarding the use of the HTQ symptom scales with specific cultural or linguistic groups. However, the moderating effects of cultural variables do provide some insight on this matter and can help clinicians as they weigh the strengths and weaknesses of this measure. Clinicians can also refer to the primary studies included in the meta-analyses for more specific psychometric information relevant to the cultural or linguistic group of interest. Clinicians are also encouraged to weigh the costs and benefits of using the HTQ as compared to other PTSD measures that have been validated for use with individuals who have similar characteristics to the client in question. For instance, if another PTSD measure has more sound psychometric evidence for a specific cultural/linguistic group than the HTQ, the other measure would be more appropriate to use.

Furthermore, because of the overlapping symptoms and high rates of comorbidity with anxiety and mood disorders, it is recommended that clinicians include the assessment of anxiety and mood-related symptoms in their differential diagnosis of PTSD. As the results from Study 2 also highlighted an important association between the HTQ symptom scales and somatization

measure, a further investigation of somatization symptoms may also be appropriate. Several other factors need to be considered in the assessment of PTSD in diverse cultures, including the evaluation of local idioms of distress, other culture-bound syndromes and culture-specific response styles.

In terms of the specific uses of the HTQ-symptom scales, in addition to diagnostic purposes, these scales may be particularly useful for the monitoring of symptoms, as they are relatively short and easy to administer. However, future studies are needed to evaluate these scales' sensitivity to change, by administering the scales at different time points and compare the scores obtained with another PTSD instrument shown to be sensitive to change, such as the CAPS (Keane et al., 2008). Although there was insufficient information available to evaluate the revised version of the HTQ (HTQ-R) in the current studies, the addition of items that assess daily functioning in this newer version may provide relevant information when monitoring the progress of treatment.

Conclusion

Common knowledge among psychologists dictates that good assessment practices include ensuring that an assessment instrument has good psychometric properties. This is true, however evidence-based assessment (EBA) goes far beyond this. In addition to using psychometrically sound measures the overall assessment process entails a complex decision-making strategy that involves a constant back and forth between the formulation and testing of hypotheses, which is often based on incomplete or contradicting information (Hunsley & Mash, 2007). Strategies to help clinicians and researchers make informed and practical decisions that are grounded in EBA practices are available in the literature (e.g., Christon, McLeod, & Jensen-

Doss, 2015). However useful these strategies are, EBA is in continuous development and is not without its struggles.

The current studies highlighted challenges in the development of EBA described by Hunsley and Mash (2007). These difficulties included: (a) defining psychometric adequacy; (b) addressing comorbidity; and (c) addressing diversity. The challenges in defining psychometric adequacy were particularly evident in Study 2 where the absence of benchmarks for validity coefficients made it difficult to fully evaluate the construct validity of the HTQ symptom scales. Further research is thus needed to establish more specific and quantitative guidelines related to the evaluation of convergent and discriminant validity. The difficulties in addressing comorbidity were also highlighted in Study 2 where the ability of the HTQ symptom scales to discriminate PTSD from other constructs, such as mood and anxiety disorders, was compromised. This issue has been an ongoing debate in the literature where some researchers argue that a dimensional or transdiagnostic approach to PTSD diagnosis may be more appropriate than conceptualizing the disorder as a discrete category (e.g., Elhai et al., 2015). Because there is evidence for both arguments, further data are needed to address the issue. Finally, addressing diversity was the most salient challenge of the three and was also the connecting thread throughout this dissertation. For instance, questions regarding the cross-cultural applicability of the HTQ trauma symptom scales (and consequently the cross-cultural applicability of the PTSD construct) arose as significant cultural variations were found in both the reliability and validity properties of the scales. Other factors pertaining to diversity such as age and gender also had moderating effects in both studies.

Despite the need for further investigation into the psychometric properties of the HTQ symptom scales, the current dissertation has shed some light on the overall psychometric strength

of this measure. Researchers and clinicians can use the information provided in both studies to help them decide whether to use this instrument for the assessment of PTSD with a specific client or sample. Potential users of the HTQ are also encouraged to complement their knowledge base with EBA guidelines to establish whether the HTQ is appropriate as is, or needs to be adapted. If an adaptation is required, researchers need to refer to best practice cross-cultural adaptation guidelines to help them in this process. In the end, following established EBA guidelines while using the HTQ will not only help maximize the accuracy of the assessment process, but also lead towards the ultimate goal of providing the best (and most deserving) care for survivors of trauma.

References (General Introduction and Discussion)

Aboraya, A. (2009). Use of structured interviews by psychiatrists in real clinical settings: Results of an open-question survey. *Psychiatry, 6*(6), 24-28.

Alcántara, C., Casement, D. M., & Lewis-Fernández, R. (2013). Conditional risk for PTSD among Latinos: A systematic review of racial/ethnic differences and sociocultural explanations. *Clinical Psychology Review, 33,* 107-119. doi: 10.1016/j.cpr.2012.10.005.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

American Psychiatric Association (APA). (1952). *Diagnostic and statistical manual of mental disorders* (1st ed.). Washington, DC: Author.

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. revised). Washington, DC: Author.

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

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychological Association Presidential Task Force on Evidence Based Practice (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271-285. doi: 10.1037/0003-066X.61.4.271

American Psychological Association Publications and Communications Board Working Group on JARS (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63,* 839-851. doi: 10.1037/0003-066X.63.9.839

Antony, M. M., & Rowa, K. (2005). Evidence-based assessment of anxiety disorders in adults. *Psychological Assessment, 17*, 256-266. doi: 10.1037/1040-3590.17.3.256

Appelbaum, M., Cooper, H, Maxwell, S., Stone, A., & Sher, K. J. (2008). Reporting standards in research in psychology: Why do we need them? What might they be? *American Psychologist, 63,* 839-851. doi: 10.1037/0003-066X.63.9.839

Armour, C., Műllerová, J., & Elhai, J. D. (2016). A systematic literature review of PTSD's latent structure in the Diagnostic and Statistical Manual of Mental Disorders: DSM-IV to DSM-5. *Clinical Psychology Review, 44*, 60-74. doi: 10.1016/j.cpr.2015.12.003

Arnetz, B. B., Broadbridge, C. L., Jamil, H., Lumley, M. A., Pole, N., Barkho, E., … Arnetz, J. E. (2014). Specific trauma subtypes improve the predictive validity of the Harvard Trauma Questionnaire in Iraqi refugees. *Journal of Immigrant and Minority Health, 16*, 1055-1061. doi: 10.1007/s10903-014-9995-9

Bahari, R., Alwi, M. N. M., Ahmad, M. R., & Saiboon, I. M. (2015). Translation and validation of the Malay Posttraumatic Stress Disorder Checklist for Civilians. *ASEAN Journal of Psychiatry, 16*(12), 203-211.

Barnes, L. L. B., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement, 62*, 603-618. doi: 10.1177/0013164402062004005

Baschnagel, J. S., O'Connor, R. M., Colder, C. R., & Hawk, L. W. (2005). Factor structure of posttraumatic stress among Western New York undergraduates following the September 11th terrorist attack on the World Trade Center. *Journal of Traumatic Stress, 18*, 677-684. doi: 10.1002/jts.20076

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Beckham, J. C., Moore, S. D., Feldman, M. E., Hertzberg, M. A., Kirby, A. C., & Fairbank, J. A. (1998). Health status, somatization, and severity of posttraumatic stress disorder in Vietnam combat veterans with posttraumatic stress disorder. *American Journal of Psychiatry, 155*, 1565-1569. doi: 10.1176/ajp.155.11.1565

Berry, J. W. (1989). Imposed etics-emics-derived etics: The operationalization of a compelling idea. *International Journal of Psychology, 24*, 721–735. doi: 10.1080/00207598908247841

Bickman, L., Nurcombe, B., Townsend, C., Belle, M., Schut, J., & Karver, M. (1999). *Consumer measurement systems in child and adolescent mental health*. Canberra, Australia: Department of Health and Family Services.

Blake, D. D., Weathers, F., Nagy, L. M., Kaloupek, D. G., Klauminzer, G., Charney, D. S., & Keane, T. M. (1990). A clinician rating scale for assessing current and lifetime PTSD: The CAPS-1. *Behavior Therapist, 13*, 187-188.

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., &

Keane, T. M. (2016). Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition (PCL-5) in veterans. *Psychological Assessment*, 28, 1379-1391. doi: 10.1037/pas0000254

Brackbill, R. M., Hadler, J. L., DiGrande, L., Ekenga, C.C., Farfel, M.R., Friedman S., ….Thorpe, L.E. (2009). Asthma and posttraumatic stress symptoms 5 to 6 years following exposure to the World Trade Center terrorist attack. *Journal of the American Medical Association*, 302, 502–516. doi: 10.1001/jama.2009.1121

Bracken, P. J., Giller, J. E., & Summerfield, D. (1995). Psychological responses to war and atrocity: the limitations of current concepts. *Social Science & Medicine, 40,* 1073-1082. doi: 10.1016/0277-9536(94)00181-R

Breslau, N., Davis, G., Andreski, P., & Peterson, E. (1991). Traumatic events and posttraumatic stress disorder in an urban population of young adults. *Archives of General Psychiatry, 48*, 216-222. doi:10.1001/archpsyc.1991.01810270028003

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: The 1996 Detroit Area Survey of Trauma. *Archives of General Psychiatry, 55,* 626-631. doi: 10.1001/archpsyc.55.7.626

Breslau, N., Troost, J. P., Bohnert, K., & Luo, Z. (2013). Influence of predispositions on post-traumatic stress disorder: does it vary by trauma severity? *Psychological Medicine, 43,* 381-390. doi: 10.1017/S0033291712001195

Brewin, C. R., Andrews, B., & Rose, S. (2003). Diagnostic overlap between acute stress disorder and PTSD in victims of violent crime. *The American Journal of Psychiatry, 160*, 783-785. doi: 10.1176/appi.ajp.160.4.783

Briere, J. (1995). *Trauma Symptom Inventory: Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185-216. doi: 10.1177/135910457000100301

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137-164). Thousand Oaks, CA: Sage Publications.

Calhoun, P. S., Hertzberg, J. S., Kirby, A. C., Dennis, M. F., Hair, L. P., Dedert, E. A., & Beckham, J. C. (2012). The effect of draft DSM-V criteria on posttraumatic stress disorder prevalence. *Depression and Anxiety, 29,* 1032-1042. doi: 10.1002/da.22012

Canadian Psychological Association Task Force on Evidence-Based Practice of Psychological Treatments (2012). *Evidence-based practice of psychological treatments: A Canadian perspective.* Retrieved from Canadian Psychological Association website: http://www.cpa.ca/docs/File/Practice/Report_of_the_EBP_Task_Force_FINAL_Board_Approved_2012.pdf

Carlson, E. B., & Rosser-Hogan, R. (1991). Trauma experiences, posttraumatic stress, dissociation, and depression in Cambodian refugees. *American Journal of Psychiatry, 148*, 1548-1551. doi: 10.1176/ajp.148.11.1548

Charney, M. E., & Keane, T. M. (2007). Psychometric analyses of the Clinician-Administered PTSD Scale (CAPS) – Bosnian translation. *Cultural Diversity & Ethnic Minority Psychology, 13,* 161-168. doi: 10.1037/1099-9809.13.2.161

Chhim, S. (2012). Baksbat (broken courage): The development and validation of the inventory to measure baksbat, a Cambodian trauma-based cultural syndrome of distress. *Culture,*

Medicine and Psychiatry, 36, 640-659. doi: 10.1007/s11013-012-9279-6

Christon, L. M., McLeod, B. D., & Jensen-Doss, A. (2015). Evidence-based assessment meets evidence-based treatment: An approach to science-informed case conceptualization. *Cognitive and Behavioral Practice, 22*, 36-48. doi: 10.1016/j.cbpra.2013.12.004

Chun, K. M., Organista, P. B., & Marín G. (Eds.). (2003). *Acculturation: Advances in theory, measurement, and applied research.* Washington, DC: American Psychological Association.

Churchill, G. A., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research, 21*(4), 360-374.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi: 10.1037/0033-2909.112.1.155

Cohen, J. A., & the Work Group on Quality Issues (1998). Practice parameters for the assessment and treatment of children and adolescents with posttraumatic stress disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*, 4S-26S. doi: 10.1097/00004583-199810001-00002

Contractor, A. A., Durham, T. A., Brennan, J. A., Armour, C., Wutrick, H. R., Frueh, B. C., & Elhai, J. D. (2014). DSM-5 PTSD's symptom dimensions and relations with major depression's symptom dimensions in a primary care sample. *Psychiatry Research, 215*, 146-153. doi: 10.1016/j.psychres.2013.10.015

Cooper, H. M. (2011). *Reporting research in psychology: How to meet journal article reporting standards* (6th ed.). Washington, DC: American Psychological Association.

Cox, B. J., Clara, I. P., & Enns, M. W. (2002). Posttraumatic stress disorder and the structure of common mental disorders. *Depression & Anxiety, 15*, 168-171. doi: 10.1002/da.10052

Creamer, M., & Parslow, R. (2008). Trauma exposure and posttraumatic stress disorder in the elderly: a community prevalence study. *The American Journal of Geriatric Psychiatry, 16*, 853-856. doi: 10.1097/01.JGP.0000310785.36837.85

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. doi:10.1007/BF02310555

da Mota Falcão, D., Ciconelli, R. M., & Ferraz, M. B. (2003). Translation and cultural adaptation of quality of life questionnaires: An evaluation of methodology. *The Journal of Rheumatology, 30*(2), 379-385.

de Fouchier, C., Blanchet, A., Hopkins, W., Bui, E., Ait-Aoudia, M., & Jehel, L. (2012). Validation of a French adaptation of the Harvard Trauma Questionnaire among torture survivors from sub-Saharan African countries. *European Journal of Psychotraumatology, 3*. doi: 10.3402/ejpt.v3i0.19225

de Jong, J. T., Komproe, I. H., Van Ommeren, M., El Masri, M., Araya, M., Khaled, N., … Somasundaram, D. (2001). Lifetime events and posttraumatic stress disorder in 4 postconflict settings. *Journal of the American Medical Association, 286*, 555-562. doi: 10.1001/jama.286.5.555

Dermichyan, A., Goenjian, A. K., & Khachadourian, V. (2015). Factor structure and psychometric properties of the Posttraumatic Stress Disorder (PTSD) Checklist and DSM-5 PTSD symptom set in a long-term post-earthquake cohort in Armenia. *Assessment, 22*, 594-606. doi: 10.1177/1073191114555523

Ditlevsen, D. N., & Elklit, A. (2010). The combined effect of gender and age on posttraumatic stress disorder: Do men and women show differences in the lifespan distribution of the disorder? *Annals of General Psychiatry, 9*. doi: 10.1186/1744-859X-9-32

Dolnicar, S., & Grün, B. (2007). Cross-cultural differences in survey response patterns. *International Marketing Review, 24,* 127-143. doi: 10.1108/02651330710741785

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*, 399-412. doi:10.1111/bjop.12046

Dye, D. A., Reck, M., & McDaniel, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment, 1*(3), 153-157.

Elhai, J. D., Contractor, A. A., Tamburrino, M., Fine, T. H., Cohen, G., Shirley, E., …, Galea, S. (2015). Structural relations between DSM-5 PTSD and major depression symptoms in military soldiers. *Journal of Affective Disorders, 175,* 373-378. doi: 10.1016/j.jad.2015.01.034

Elhai, J. D., Gray, M. J., Kashdan, T. B., & Franklin, C. L. (2005). Which instruments are most commonly used to assess traumatic event exposure and posttraumatic effects? A survey of traumatic stress professionals. *Journal of Traumatic Stress, 18,* 541-545. doi: 10.1002/jts.20062

Elklit, A., & Shevlin, M. (2007). The structure of PTSD symptoms: A test of alternative models using confirmatory factor analysis. *British Journal of Clinical Psychology, 46*, 299-313. doi: 10.1348/014466506X171540

Elsass, P. (2001). Individual and collective traumatic memories: A qualitative study of posttraumatic stress disorder symptoms in two Latin American localities. *Transcultural Psychiatry, 38*, 306-316. doi: 10.1177/136346150103800302

Epstein, J., Osborne, R. H., Elsworth, G. R., Beaton, D. E., & Guillemin, F. (2013). Cross-cultural adaption of the Health Education Impact Questionnaire: experimental study

showed expert committee, not back-translation, added value. *Journal of Clinical Epidemiology, 68,* 360-369. doi: 10.1016/j.jclinepi.2013.07.013

Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology, 68,* 435-441. doi: 10.1016/j.jclinepi.2014.11.021

Eshun, S., & Gurung, R. A. (Eds.). (2009). *Culture and mental health: Sociocultural influences, theory and practice.* Malden, MA: Blackwell Publishing Ltd.

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

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured Clinical Interview for DSM-IV Axis I disorders, Clinician version.* Washington, DC: American Psychiatric Press, Inc.

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2016). *Structured Clinical Interview for DSM-5 Disorders – Clinician version.* Arlington, WA: American Psychiatric Association Publishing.

Flaherty, J. A., Gaviria, F. M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J. A., & Birz, S. (1988). Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disease, 176,* 257-263. doi: 10.1097/00005053-198805000-00001

Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment, 9,* 445– 451. doi: 10.1037/1040-3590.9.4.445

Friedman, M. J. (2013). Finalizing PTSD in DSM-5: Getting here from there and where to go next. *Journal of Traumatic Stress, 26,* 548-556. doi: 10.1002/jts.21840

Gagnon, A. J., Tuck, J., & Barkun, L. (2004). A systematic review of questionnaires measuring the health of resettling refugee women. *Health Care for Women International, 25*, 111-149. doi: 10.1080/07399330490267503

Gentes, E. L., Dennis, P. A., Kimbrel, N. A., Rissling, M. B., Beckham, J. C., VA Mid-Atlantic MIRECC Workgroup, & Calhoun, P. S. (2014). DSM-5 posttraumatic stress disorder: Factor structure and rates of diagnosis. *Journal of Psychiatric Research, 59,* 60-67. doi: 10.1016/j.jpsychires.2014.08.014

Gorst-Unsworth, C., & Goldenberg, E. (1998). Psychological sequelae of torture and organised violence suffered by refugees from Iraq. Trauma-related factors compared with social factors in exile. *The British Journal of Psychiatry, 172,* 90-94. doi: 10.1192/bjp.172.1.90

Graham, J. M., Yenling, J. L., & Jeziorski, J. L. (2006). The Dyadic Adjustment Scale: A reliability generalization meta-analysis. *Journal of Marriage and Family, 68,* 701-717. doi: 10.1111/j.1741-3737.2006.00284.x

Grubaugh, A. L., Long, M. E., Elhai, J. D., Frueh, B. C., & Magruder, K. M. (2010). An examination of the construct validity of posttraumatic stress disorder with veterans using a revised criterion set. *Behavior, Research and Therapy. 48*, 909-914. doi: 10.1016/j.brat.2010.05.019

Guajardo, M. G. U., Slewa-Younan, S., Smith, M., Eagar, S., & Stone, G. (2016). Psychological distress is influenced by length of stay in resettled Iraqi refugees in Australia. *International Journal of Mental Health Systems, 10*, 4. doi: 10.1186/s13033-016-0036-z

Hagengimana, A., & Hinton, D. (2009). Ihahamuka , a Rwandan syndrome of response to the

genocide: Blocked flow, spirit assault, and shortness of breath. In D. E. Hinton & B. J. Good (Eds.), *Culture and panic disorder* (pp. 205-229). Stanford, CA: Stanford University Press.

Harvard Program in Refugee Trauma (n.d.). Harvard Trauma Questionnaire: Kurdish version.

Harvard Program in Refugee Trauma (n.d.). Harvard Trauma Questionnaire: Russian version.

Harvard Program in Refugee Trauma (1996). Harvard Trauma Questionnaire: Japanese version.

Harvard Program in Refugee Trauma (1998). Harvard Trauma Questionnaire: Bosnian version.

Harvard Program in Refugee Trauma (1998). Harvard Trauma Questionnaire: Croatian version.

Harvard Program in Refugee Trauma (1998). Harvard Trauma Questionnaire: revised English version.

Harvard Program in Refugee Trauma (2000). Harvard Trauma Questionnaire: revised Cambodian version.

Harvard Program in Refugee Trauma (2000). Harvard Trauma Questionnaire: Peruvian version.

Harvard Program in Refugee Trauma (2000). Harvard Trauma Questionnaire: Spanish version.

Harvard Program in Refugee Trauma (2006). Harvard Trauma Questionnaire: Arabic version.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross-Cultural Management, 6*, 243-266. doi: 10.1177/1470595806066332

Hauff, E., & Vaglum, P. (1994). Chronic posttraumatic stress disorder in Vietnamese refugees. A prospective community study of prevalence, course, psychopathology, and stressors. *Journal of Nervous and Mental Disease, 182*(2), 85-90.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment,*

7, 238-247. doi: 10.1037/1040-3590.7.3.238

Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2011). *Scientific foundations of clinical assessment.* New York, NY: Routledge.

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, 2*. Retrieved from http://scholarworks.gvsu.edu/cgi/viewcontent.cgi?article=1111&context=orpc

Hemphill, J. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58*, 78-79. doi: 10.1037/0003-066X.58.1.78

Henchy, A. M. (2012). *Review and evaluation of reliability generalization research* (Unpublished doctoral dissertation). University of Kentucky, Lexington, Kentucky.

Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement, 61*, 404-420. doi: 10.1177/00131640121971284

Hinton, D. E., Chhean, D., Pich, V., Um, K., Fama, J. M., & Pollack, M. H. (2006). Neck-focused panic attacks among Cambodian refugees: A logistic and linear regression analysis. *Journal of Anxiety Disorders, 20*, 119-138. doi: 10.1016/j.janxdis.2005.02.001

Hinton, D. E., Hinton, A., Chhean, D., Pich, V., Loeum, J. R., & Pollack, M. H. (2009). Nightmares among Cambodian refugees: The breaching of concentric ontological security. *Culture, Medicine, and Psychiatry, 33*, 219-265. doi: 10.1007/s11013-009-9131-9

Hinton, D. E., Hofmann, S. G., Pitman, R. K., Pollack, M. H., & Barlow, D. H. (2008). The panic attack-PTSD model: Applicability to orthostatic panic among Cambodian refugees. *Cognitive Behaviour Therapy, 37*, 101-116. doi: 10.1080/16506070801969062

Hinton, D. E., & Lewis-Fernández, R. (2010). Idioms of distress among trauma survivors: Subtypes and clinical utility. *Culture, Medicine, and Psychiatry, 34*, 209–218. doi: 10.1007/s11013-010-9175-x

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

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the Mind*, (3rd ed.). New York: McGraw-Hill.

Hollifield, M., Hewage, C., Gunawardena, C. N., Kodituwakku, P., Bopagoda, K., & Weerarathnege, K. (2008). Symptoms and coping in Sri Lanka 20-21 months after the 2004 tsunami. *The British Journal of Psychiatry, 192*, 39-44. doi: 10.1192/bjp.bp.107.038422

Hollifield, M., Warner, T. D., Lian, N., Krakow, B., Jenkins, J. H., Kesler, J., … Westermeyer, J. (2002). Measuring trauma and health status in refugees: A critical review. *Journal of the American Medical Association, 288*, 611-621. doi: 10.1001/jama.288.5.611

Howell, R. T., & Shields, A. L. (2008). The file drawer problem in reliability generalization: A strategy to compute a fail-safe *N* with reliability coefficients. *Educational and Psychological Measurement, 68*, 120-128. doi:10.1177/0013164407301528

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3*, 29-51. doi: 10.1146/annurev.clinpsy.3.022806.091419

Hunsley, J., & Mash, E. J. (2008). Developing criteria for evidence-based assessment: An introduction to assessments that work. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 3-14). New York: Oxford University Press.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15*, 446-455. doi: 10.1037/1040-3590.15.4.446

Ichikawa, M., Nakahara, S., & Wakai, S. (2006). Effect of post-migration detention on mental health among Afghan asylum seekers in Japan. *Australian and New Zealand Journal of Psychiatry, 40*, 341-346. doi: 10.1080/j.1440-1614.2006.01800.x

International Test Commission (2005). *International Guidelines on Test Adaptation*. Retrieved from: http://www.intestcom.org

Jakobsen, M., Thoresen, S., & Johansen, L. E. E. (2011). The validity of screening for post-traumatic stress disorder and other mental health problems among asylum seekers from different countries. *Journal of Refugee Studies, 24*, 171-186. doi: 10.1093/jrs/feq053

Jayawickreme, N., Jayawickreme, E., & Foa, E. B. (2012). Using the individualism-collectivism construct to understand cultural differences in posttraumatic stress disorder. In K. M. Gow, & M. J. Celinski (Eds.), *Wayfinding through life's challenges: Coping and survival*. New York: Nova Science Publishers.

Jenkins, J. H., & Valiente, M. (1994). Bodily transactions of the passions: El calor (the heat) among Salvadoran women. In T. Csordas (Ed.), *The body as existential ground: Studies in culture, self, and experience* (pp. 163-182). Cambridge, UK: Cambridge University Press.

Jobson, L., & O'Kearney, R. (2008). Cultural differences in personal identity in posttraumatic stress disorder. *British Journal of Clinical Psychology, 47*, 95-109. doi: 10.1348/014466507X235953

Jones, E., Vermaas, R. H., McCartney, H., Beech, C., Palmer, I., Hyams, K., & Wessely, S. (2003). Flashbacks and posttraumatic stress disorder: The genesis of a 20th-century diagnosis. *British Journal of Psychiatry, 182,* 158–163. doi: 10.1192/bjp.182.2.158

Kawakami, N., Takeshima, T., Ono, Y., Uda, H., Hata, Y., Nakane, Y., … Kikkawa, T. (2005). Twelve-month prevalence, severity, and treatment of common mental disorders in communities in Japan: preliminary finding from the World Mental Health Japan Survey 2002–2003. *Psychiatry and Clinical Neurosciences, 59*, 441-452. doi: 10.1111/j.1440-1819.2005.01397.x

Keane, T. M. (1995). Guidelines for the forensic psychological assessment of posttraumatic stress disorder claimants. In R. I. Simon (Ed.), *Posttraumatic stress disorder in litigation: Guidelines for forensic assessment* (pp. 99-115). Washington, DC: American Psychiatric Press, Inc.

Keane, T. M., Buckley, T., & Miller, M. (2003). Guidelines for the forensic psychological assessment of posttraumatic stress disorder claimants. In R. I. Simon (Ed.), *Posttraumatic stress disorder in litigation: Guidelines for forensic assessment* (2nd ed., pp. 119-140). Washington, DC: American Psychiatric Association Press, Inc.

Keane, T. M., Fairbank, J. A., Caddell, J. M., Zimering, R. T., & Bender, M. E. (1985). A behavioral approach to assessing and treating posttraumatic stress disorder in Vietnam veterans. In C. R. Figley (Ed.), *Trauma and its wake* (pp. 257–294). New York: Brunner/Mazel.

Keane, T. M., Kaloupek, D. G., & Weathers, F. W. (1996). Ethnocultural considerations in the assessment of PTSD. In A. J. Marsella, M. J. Friedman, E. T. Gerrity & R. M. Scurfield (Eds.), *Ethnocultural aspects of posttraumatic stress disorder: Issues, research, and*

clinical applications (pp. 183-205). Washington, DC: American Psychological Association.

Keane, T. M., Kolb, L. C., Kaloupek, D. G., Orr, S. P., Blanchard, E. B., Thomas, R. G., … Lavori, P. W. (1998). Utility of psychophysiological measurement in the diagnosis of posttraumatic stress disorder: results from a Department of Veterans Affairs Cooperative Study. *Journal of Consulting and Clinical Psychology, 66,* 914-923. doi: 10.1037/0022-006X.66.6.914

Keane, T. M., Silberbogen, A. K., & Weierich, M. R. (2008). Posttraumatic stress disorder. In J. Hunsley & E.J. Mash (Eds.), *A guide to assessments that work* (pp. 293-315). New York: Oxford University Press.

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry, 52*, 1048-1060. doi: 10.1002/1099-1298(200011/12)

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., Nelson, C. B., & Breslau, N. (1999). Epidemiological risk factors for trauma and PTSD. In R. Yehuda (Ed.), *Risk factors for posttraumatic stress disorder* (pp. 23-60). Washington, DC: American Psychiatric Press.

Kilpatrick, D. G., Resnick, H. S., Milanak, M. E., Miller, M. W., Keyes, K. M., & Friedman, M. J. (2013). National estimates of exposure to traumatic events and PTSD prevalence using DSM-IV and DSM-5 criteria. *Journal of Traumatic Stress, 26,* 537-547. doi: 10.1002/jts.21848

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments in research. *American Journal of Health-System Pharmacists, 65,* 2276-2284. doi: 10.2146/ajhp070364

King, D. W., Leskin, G. A., King, L. A., & Weathers, F. W. (1998). Confirmatory factor analysis of the clinician-administered PTSD scale: evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment, 10,* 90-96. doi: 10.1037/1040-3590.10.2.90

Kirmayer, L. J., & Young, A. (1998). Culture and somatization: Clinical, epidemiological, and ethnographic perspectives. *Psychosomatic Medicine, 60*(4), 420-430.

Kleijn, W. C., Hovens, J. E. & Rodenburg, J. J. (2001). Posttraumatic stress symptoms in refugees: Assessments with the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist-25 in different languages. *Psychological Reports, 88,* 527-532. doi: 10.2466/PR0.88.2.527-532

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., … Kessler, R. C. (2017). Posttraumatic stress disorder in the World Mental Health Surveys, *Psychological Medicine,* 1-15. doi: 10.1017/S0033291717000708

Krause, E. D., Kaltman, S., Goodman, L. A., & Dutton, M. A. (2007). Longitudinal factor structure of posttraumatic stress disorder symptoms related to intimate partner violence. *Psychological Assessment, 19,* 165-175. doi: 10.1037/1040-3590.19.2.165

Laban, C. J., Gernaat, H. B., Komproe, I. H., van der Tweel, I., & de Jong, J. T. (2005). Postmigration living problems and common psychiatric disorders in Iraqi asylum seekers in the Netherlands. *Journal of Nervous and Mental Disease, 193,* 825-832. doi: 00005053-200412000-00007

Lange, A., Rietdijk, D., Hudcovicova, M., van de Ven, J. P., Schrieken, B., & Emmelkamp, P. M. (2003). Interapy: A controlled randomized trial of the standardized treatment of

posttraumatic stress through the internet. *Journal of Consulting and Clinical Psychology, 71*, 901–909. doi: 10.1037/0022-006X.71.5.901

Lauterbach, D., Vrana, S., King, D. W., & King, L. A. (1997). Psychometric properties of the Civilian Version of the Mississippi PTSD Scale. *Journal of Traumatic Stress, 10*, 499-513. doi: 10.1002/jts.2490100313

Leiknes, K. A., Finset, A., Moum, T., & Sandanger, I. (2007). Course and predictors of medically unexplained pain symptoms in the general population. *Journal of Psychosomatic Research, 62*, 119-128. doi: 10.1016/j.jpsychores.2006.08.009

Lewis-Fernández, R., Hinton, D. E., & Marques, L. (2014). Culture and PTSD. In M. J. Friedman, T. M. Keane & P. A. Resick (Eds.) *Handbook of PTSD* (2nd ed., pp. 522-539). New York, NY: Guilford Press.

López- López, J. A., Botella, J., Sánchez-Meca, J. & Marín-Martínez, F. (2013). Alternatives for mixed-effects meta-regression models in the reliability generalization approach: A simulation study. *Journal of Educational and Behavioral Statistics, 38*, 443-469. doi: 10.3102/1076998612466142

Maneesriwongul, W. & Dixon, J. K. (2004) Instrument translation process: A methods review. *Journal of Advanced Nursing, 48*, 175-186. doi: 10.1111/j.1365-2648.2004.03185.x

Marsella, A. J., Friedman, M. J., Gerrity, E. T. & Scurfield, R. M. (1996). *Ethnocultural aspects of posttraumatic stress disorder: Issues, research, and clinical applications*. Washington, DC: American Psychological Association.

Matthieu, M., & Ivanoff, A. (2006). Treatment of human-caused trauma: Attrition in the adult outcomes research. *Journal of Interpersonal Violence, 2*, 1654–1664. doi: 10.1177/0886260506294243

McCall, G., & Resick, P. (2003). A pilot study of PTSD symptoms among Kalahari Bushmen. *Journal of Traumatic Stress, 16,* 445–450. doi: 10.1023/A:1025702326392

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

Medina-Mora, M., Borges, G., Lara, C., Benjet, C., Blanco, J., Fleiz, C., … Zambrano, J. (2005). Prevalence, service use, and demographic correlates of 12-month DSM-IV psychiatric disorders in Mexico: Results from the Mexican National Comorbidity Survey. *Psychological Medicine, 35*, 1-11. doi: 10.1017/S0033291705005672

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' response and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749. doi: 10.1037/0003-066X.50.9.741

Michalopoulos, L. M., Unick, G. J., Haroz, E. E., Bass, J., Murray, L. K., & Bolton, P. A. (2015). Exploring the fit of western PTSD models across three non-western low- and middle-income countries. *Traumatology, 21*, 55-63. doi: 10.1037/trm0000020

Miles, J. N., Marshall, G. N., & Schell, T. L. (2008). Spanish and English versions of the PTSD Checklist-Civilian version (PCL-C): Testing for differential item functioning. *Journal of Traumatic Stress, 21*, 369-376. doi: 10.1002/jts.20349

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*(1), 111-121.

Miller, M. W., Fogler, J. M., Wolf, E. J., Kaloupek, D. G., & Keane, T. M. (2008). The internalizing and externalizing structure of psychiatric comorbidity in combat veterans. J*ournal of Traumatic Stress, 21*, 58–65. doi: 10.1002/jts.20303

Miller, M. W., Kaloupek, D. G., Dillon, A. L., & Keane, T. M. (2004). Externalizing and

internalizing subtypes of combat-related PTSD: A replication and extension using the PSY-5 scales. *Journal of Abnormal Psychology, 113*, 636–645. doi: 10.1037/0021-843X.113.4.636

Miller, K. E., Omidian, P., Kulkarni, M., Yaqubi, A., Daudzai, H., & Rasmussen, A. (2009). The validity and clinical utility of posttraumatic stress disorder in Afghanistan. *Transcultural Psychiatry, 46*, 219–237. doi: 10.1177/1363461509105813

Miller, K. E., Omidian, P. A., Quraishy, A. S., Quraishy, N., Nasiry, M. N., Nasiry, S., … Yaqubi, A. A. (2006). The Afghan Symptom Checklist: A culturally grounded approach to mental health assessment in a conflict zone. *American Journal of Orthopsychiatry, 76*, 423–433. doi: 10.1037/0002-9432.76.4.423

Mills, J., & Clark, M. S. (1982). Communal and exchange relationships. *Review of Personality and Social Psychology, 3*, 121-144.

Mollica, R. F., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., & Lavelle, J. (1992). The Harvard Trauma Questionnaire: Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *Journal of Nervous and Mental Disease, 180*, 111-116. doi: 10.1097/00005053-199202000-00008

Mollica, R. F., Caspi-Yavin, Y., Lavelle, J., Tor, S., Yang, T., Chan, S., ... de Marneffe, D. (1994). *The Harvard Trauma (HTQ) Manual: Cambodian, Laotian, and Vietnamese versions* (Unpublished manual). Cambridge, MA.

Mollica, R. F., McDonald, L. S., Massagli, M. P., & Silove, D. M. (2004). *Measuring trauma, measuring torture: Instructions and guidance on the utilization of the Harvard Program in Refugee Trauma's versions of the Hopkins Symptom Checklist-25 (HSCL-25) and the Harvard Trauma Questionnaire (HTQ)* (Unpublished manual). Cambridge, MA.

Mollica, R. F., McInness, K., Poole, C., & Tor, S. (1998). Dose–effect relationships of trauma to symptoms of depression and post-traumatic stress disorder among Cambodian survivors of mass violence. *British Journal of Psychiatry, 173*, 482-488. doi: 10.1192/bjp.173.6.482

Mollica, R. F., Wyshak, G., de Marneffe, D., & Lavelle, J. (1987). Indochinese versions of the Hopkins Symptom Checklist-25: A screening instrument for the psychiatric care of refugees. *American Journal of Psychiatry, 144*, 497-500. doi: 10.1176/ajp.144.4.497

Murphy, H. B. M. (1982) *Comparative psychiatry: The International and intercultural distribution of mental illness*. Berlin, Germany: Springer-Verlag.

Nakeyar, C., & Frewen, P. A. (2016). Evidence-based care for Iraqi, Kurdish, and Syrian asylum seekers and refugees of the Syrian civil war: A systematic review. *Canadian Psychology, 57*, 233-245. doi: 10.1037/cap0000067

Nock, M. K., Hwang, I., Sampson, N. A., & Kessler, R. C. (2010). Mental disorders, comorbidity and suicidal behavior: Results from the National Comorbidity Survey Replication. *Molecular Psychiatry, 15*, 868–876. doi:10.1038/mp.2009.29

North, C. S., Kawasaki, A., Spitznagel, E. L., & Hong, B. A. (2004). The course of PTSD, major depression, substance abuse, and somatization after a natural disaster. *Journal of Nervous and Mental Disease, 192*, 823-829. doi: 10.1097/01.nmd.0000146911.52616.22

Norris, A. E., & Aroian, K. J. (2008). Avoidance symptoms and assessment of posttraumatic stress disorder in Arab immigrant women. *Journal of Traumatic Stress, 21*, 471-478. doi: 10.1002/jts.20363

Norris, F. H. (1992). Epidemiology of trauma: frequency and impact of different potentially traumatic events on different demographic groups. *Journal of Consulting and Clinical Psychology, 60,* 409-418. doi: 10.1037/0022-006X.60.3.409

Norris, F. H., & Slone, L. B. (2014). Epidemiology of trauma and PTSD. In M. J. Friedman, T. M. Keane & P. A. Resick (Eds.), *Handbook of PTSD: Science and practice* (2nd ed., pp. 100-120). New York: The Guilford Press.

Norris, F. H., Van Landingham, M. J., & Vu, L. (2009). PTSD in Vietnamese Americans following Hurricane Katrina: Prevalence, patterns, and predictors. *Journal of Traumatic Stress, 22,* 91-101. doi: 10.1002/jts.20389

O'Donnell, M. L., Alkemade, N., Nickerson, A., Creamer, M., McFarlane, A. C., Silove, D.,…Forbes, D. (2014). Impact of the diagnostic changes to post-traumatic stress disorder for DSM-5 and the proposed changes to ICD-11. *British Journal of Psychiatry, 205*, 230-235. doi: 10.1192/bjp.bp.113.135285

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality, 17*, S19-S38. doi: 10.1002/per.487

Orr, S. P., Metzger, L. J., Miller, M. W., & Kaloupek, D. G. (2004). Psychophysiological assessment of PTSD. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (2nd ed., pp. 289-343). New York: Guilford Press.

Palinkas, L. A., Petterson, J. S., Russell, J. C., & Downs, M. A. (2004). Ethnic differences in symptoms of posttraumatic stress after the Exxon Valdez oil spil. *Prehospital and Disaster Medicine, 19,* 102-112. doi: 10.1017/S1049023X00001552

Parry-Jones, B., & Parry-Jones, W. L. L. (1994). Posttraumatic stress disorder: Supportive evidence from and eighteenth century natural disaster. *Psychological Medicine, 24*, 15-27. doi: 10.1017/S0033291700026799

Paunović, N., & Öst, L.-G. (2005). Psychometric properties of a Swedish translation of the Clinician Administered PTSD Scale-Diagnostic version. *Journal of Traumatic Stress, 18*, 161-164. doi: 10.1002/jts.20013

Perkonigg, A., Kessler, R. C., Storz, S., & Wittchen, H.-U. (2000). Traumatic events and post-traumatic stress disorder in the community: Prevalence, risk factors and comorbidity. *Acta Psychiatrica Scandinavica, 101*, 46-59. doi: 10.1034/j.1600-0447.2000.101001046.x

Peter, J. P., & Churchill, G. A. (1986). Relationships among research design choices and psychometric properties of rating scales: A meta-analysis. *Journal of Marketing Research, 23*, 1–10. doi: 10.2307/3151771

Pietrzak, R. H., Goldstein, R. B., Southwick, S. M., & Grant, B. F. (2011). Prevalence and Axis I comorbidity of full and partial posttraumatic stress disorder in the United States: Results from Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Anxiety Disorders, 25*, 456-465. doi: 10.1016/j.janxdis.2010.11.010

Pole, N., Gone, J. P., & Kulkarni, M. (2008). Posttraumatic stress disorder among ethnoracial minorities in the United States. *Clinical Psychology: Science and Practice, 15*, 35–61. doi: 10.1111/j.1468-2850.2008.00109.x

Porter, M., & Haslam, N. (2005). Predisplacement and postdisplacement factors associated with mental health of refugees and internally displaced persons: A meta-analysis. *Journal of the American Medical Association, 294*, 602-612. doi:

10.1001/jama.294.5.602.

Rasmussen, A., Verkuilen, J., Ho, E., & Fan, Y. (2015). Posttraumatic stress disorder among refugees: Measurement invariance of Harvard Trauma Questionnaire scores across global regions and response patterns. *Psychological Assessment, 27*, 1160-1170. doi: 10.1037/pas0000115

Renner, W., Salem, I., & Ottomeyer, K. (2007). Posttraumatic stress in asylum seekers from Chechnya, Afghanistan, and West Africa: Differential findings obtained by quantitative and qualitative methods in three Austrian samples. In J. P. Wilson & C. C. Tang (Eds.), *Cross-cultural assessment of psychological trauma and PTSD* (pp. 239-278). Boston, MA: Springer.

Roberts, A. L., Gilman, S. E., Breslau, J., Breslau, N., & Koenen, K. C. (2011). Race/ethnic differences in exposure to traumatic events, development of post-traumatic stress disorder, and treatment-seeking for post-traumatic stress disorder in the United States. *Psychological Medicine, 41*, 71-83. doi: 10.1017/S0033291710000401

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F, Burke, J., … Towle, M. S. (1988). The Composite International Diagnostic Interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry, 45*, 1069-1077. doi: 10.1001/archpsyc.1988.01800360017003

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes.* San Diego, CA: Elsevier.

Ruggiero, K. J., Del Ben, K., Scotti, J. R., & Rabalais, A. E. (2003). Psychometric properties of the PTSD Checklist-Civilian version. *Journal of Traumatic Stress, 16*, 495-502. doi:

10.1023/A:1025714729117

Salkind, N. J. (2012). Tests and measurement for people who (think they) hate tests and measurement (2nd ed.). Los Angeles, CA: Sage Publications.

Sam, D. L., & Moreira, V. (2002). The mutual embeddedness of culture and mental illness. In W. J. Lonner, D. L. Dinnel, S. A. Hayes, & D. N. Sattler (Eds.), *Online readings in psychology and culture* (Unit 9, Chapter 1). Bellingham, WA: Center for Cross-Cultural Research, Western Washington University.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529-540. doi: 10.1037/0021-9010.62.5.529

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274. doi: 10.1.1.172.1733

Schnyder, U., & Moergeli, H. (2002). German version of Clinician-Administered PTSD Scale. *Journal of Traumatic Stress, 15*, 487-492. doi: 10.1023/A:1020922023090

Schwarz-Nielsen, K. H., & Elklit, A. (2009). An evaluation of the mental status of rejected asylum seekers in two Danish asylum centers. *Torture, 19*(1), 51-59.

Shoeb, M., Weinstein, H., & Mollica, R. (2007). The Harvard Trauma Questionnaire: Adapting a cross-cultural instrument for measuring torture, trauma and posttraumatic stress disorder in Iraqi refugees. *The International Journal of Social Psychiatry, 53*, 447-463. doi: 10.1177/0020764007078362

Shore, J. H., Orton, H., & Manson, S. M. (2009). Trauma-related nightmares among American Indian veterans: Views from the dream catcher. *American Indian and Alaska Native*

Mental Health Research, 16, 25-38. doi: 10.5820/aian.1601.2009.25

Silove, D., Manicavasagar, V., Mollica, R., Thai, M., Khiek, D., Lavelle, J., & Tor, S. (2007). Screening for depression and PTSD in a Cambodian population unaffected by war: Comparing the Hopkins Symptom Checklist and Harvard Trauma Questionnaire with the structured clinical interview. *Journal of Nervous and Mental Disease, 195,* 152-157. doi: 10.1097/01.nmd.0000254747.03333.70

Speroff, T., Sinnot, P. L., Marx, B., Owen, R. R., Jackson, J. C., Greevy, R., … Friedman, M. J. (2012). Impact of evidence-based standardized assessment on the disability clinical interview for diagnosis of service-connected PTSD: A cluster-randomized trial. *Journal of Traumatic Stress, 25,* 607-615. doi: 10.1002/jts.21759

Spitzer, R. L., First, M. B., & Wakefield, J. C. (2007). Saving PTSD from itself in DSM-V. *Journal of Anxiety Disorders, 21,* 233-241. doi: 10.1016/j.janxdis.2006.09.006

Steel, Z., Chey, T., Silove, D., Marmane, C., Bryant, R. A., & van Ommeren, M. (2009). Association of torture and other potentially traumatic events with mental health outcomes among populations exposed to mass conflict and displacement: A systematic review and meta-analysis. *Journal of the American Medical Association, 302,* 537-549. doi: 10.1001/jama.2009.1132

Summerfield, D. (1999). A critique of seven assumptions behind psychological trauma programmes in war-affected areas. *Social Science and Medicine, 48,* 1449–1462. doi: 10.1016/S0277-9536(98)00450-X

Summerfield, D. (2004). Cross-cultural perspectives on the medicalization of human suffering. In G. M. Rosen (Ed.), *Posttraumatic stress disorder: Issues and controversy* (pp. 233–245). Chichester, UK: John Wiley & Sons, Ltd.

Tang, S. S., & Fox, S. H. (2001). Traumatic experiences and the mental health of Senegalese refugees. *Journal of Nervous and Mental Disease, 189*, 507-512.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55. doi: 10.5116/ijme.4dfb.8dfd

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54*(4), 837-847.

Thompson, B. (1999, January). *Reliability generalization: An important meta-analysis method, because it is incorrect to say "the test is unreliable."* Paper presented at the meeting of the Southwest Educational Research Association, San Antonio, TX.

Triandis, H. C. (2001). Individualism-collectivism and personality. *Personality, 69*, 907-924. doi: 10.1111/1467-6494.696169

Trimble, M. R. (1981). *Post traumatic neuroses*. Chichester, UK: John Wiley & Sons.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*, 6-20. doi: 10.1177/0013164498058001002

Vacha-Haase, T., Kogan, L.R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement, 60*, 509–522. doi: 10.1177/00131640021970682

Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development, 44*, 159-168. doi: 10.1177/0748175611409845

Van der Velden, P. G., Grievink, L., Kleber, R. J., Drogendijk, A. N., Roskam, A. J.,

Marcelissen, F. G., … Gersons, B. P. (2006). Post-disaster mental health problems and the utilization of mental health services: A four-year longitudinal comparative study. *Administration and Policy in Mental Health, 33,* 279–288. doi: 10.1007/s10488-005-0027-x

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue européenne de psychologie appliquée, 54,* 119-135. doi: 10.1016/j.erap.2003.12.004

Vassar, M., Knaup, K.G., Hale, W., & Hale, H. (2011). A meta-analysis of coefficient alpha for the Impact of Event Scales: A reliability generalization study. *South African Journal of Psychology, 41,* 6-16. doi: 10.1177/008124631104100102

Vloeberghs, E., van der Kwaak, A., Knipscheer, J., & van den Muijsenbergh (2012). Coping and chronic psychosocial consequences of female genital mutilation in the Netherlands. *Ethnicity and Health, 17,* 677-695. doi: 10.1080/13557858.2013.771148

Wang, R., Su, J., Bi., X., Wei, Y., & Mo, L. (2012). Application of the Chinese Posttraumatic Stress Disorder Checklist to adolescent earthquake survivors in China. *Social Behavior and Personality: An International Journal, 40,* 415-423. doi: 10.2224/sbp.2012.40.3.415

Warne, R. (2008, February). *Reliability generalization (RG) analysis: The test is not reliable*. Paper presented at the meeting of the Southwest Educational Research Association, New Orleans, LA.

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

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD checklist (PCL): Reliability, validity, and diagnostic utility.* Poster presented at the 9th annual meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

White, J., Pearce, J., Morrison, S., Dunstan, F., Bisson, J. I., & Fone, D. L. (2015). Risk of post-traumatic stress disorder following traumatic events in a community sample. *Epidemiology and Psychiatric Sciences, 24*, 249-257. doi: 10.1017/S2045796014000110

Wilkinson, L., & American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604. doi:10.1037//0003-066X.54.8.594

Williams, D. R., Herman, A., Stein, D. J., Heeringa, S. G., Jackson, P. B., Moomal, H., & Kessler, R. C. (2008). Twelve-month mental disorders in South Africa: Prevalence, service use and demographic correlates in the population-based South African Stress and Health Study. *Psychological Medicine, 38*, 211-220. doi: 10.1017/S0033291707001420

Yeomans, P. D., & Foreman, E. M. (2009). Cultural factors in traumatic stress. In S. Eshun, & R. A. R. Gurung (Eds.), *Culture and mental health: Sociocultural influences, theory, and practice* (pp. 221-244). Chichester, U.K.: John Wiley & Sons Ltd.

Yufik, T., & Simms, L. J. (2010). A Meta-analytic investigation of the structure of posttraumatic stress symptoms. *Journal of Abnormal Psychology, 119*, 764-776. doi: 10.1037/a0020981

Zinbarg, R., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β and McDonald's ω_η: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123–133. doi:10.1007/s11336-003-0974-7

Appendix A

Reliability and Validity Generalizations Coding Manual

Reliability and Validity Generalizations Coding Manual

Identification information

1) Document identifier: The document identifier is comprised of two sets of numbers. The first set is the year of publication and the second set is the document number (e.g., 2010-18).

2) Sample of study: Indicate which sample of the study is used to calculate the reliability or validity coefficients. If there are coefficients for multiple sub-samples (e.g., control group, treatment group) within the same study complete separate coding sheets for each sub-sample.

Sample and sampling characteristics

3) Sample size: Code the sample size used to calculate the reliability/validity coefficients (sometimes this number is not the total sample size of the study).

4-5) Age and gender: Code age & gender provided (even if this information is only provided for the total sample and not the sample that is used for the reliability/validity coefficients).

6) Culture/Ethnicity: Indicate how the sample was described. There can be multiple cultures/ethnicities. If not specified, code as culture/ethnicity of country of study (e.g.: a study conducted in Denmark with no ethnicities/cultures specified, code "Danish"), unless it was evident that there were various cultures/ethnicities such as sample comprised of refugees.

7) Country of origin (COO): Indicate each country if there are multiple. Code "various" if there are too many to code (i.e., more than five).

8) Cultural orientation of COO: Please refer to the following website: http://geert-hofstede.com/countries.html. Select whether the country's culture is considered individualist (score > 50) or collectivist (score < 50). If the country in question is not listed on the website, select "information unavailable".

9) Country of study: Indicate each country if there are multiple. Code "various" if there are too many to code (i.e., more than five).

10) Recruitment setting: Indicate whether the sample was recruited in a mental health setting (also includes psychiatric treatment and counselling), medical treatment setting (e.g., hospital, primary care setting), community setting, or mixed sampling procedure.

11) Trauma category: Indicate which of the following traumatic events category best fits the trauma experienced by the sample (select as many as necessary): (a) war, combat or terrorism (e.g., war, civil conflict, political instability, political violence, ethnic persecution, human rights violations); (b) torture (i.e., sample was described as torture survivors [including political prisoners] – select only this category even if the sample also experienced war or combat); (c) human trafficking (e.g., sex trafficking, trafficking for labour exploitation); (d) sexual assault (e.g., sexual assault, rape, childhood sexual abuse, sexual violence); (e) nonsexual assault (e.g., robbery, crime victim, physical assault); (f) personal illness or injury (i.e., participants had an illness [e.g., cancer] or injury [e.g., whiplash from car accident]; (g) witnessing death or injury (e.g., witnessing a stabbing, witnessing a murder); (h) death, injury or illness of loved one; (i) man-made disaster or fire (e.g., explosions, house fire); (j) natural disasters (i.e., disasters from nature [e.g., tsunami, earthquake]); (k) various (i.e., participants experienced a range of traumas [e.g., probability surveys]; (l) none (e.g., control group with no prior history of trauma). Code "N/R" if the type of trauma is not indicated.

12) Forcibly displaced persons: Code whether the sample consisted of participants that were forcibly displaced (e.g., refugees, asylum seekers, internally displaced) at any point in time (i.e., participants do not need to be forcibly displaced at the time of study).

Description of the HTQ administration and adaptation procedures

13) Language of administration of the Harvard Trauma Questionnaire (HTQ): Indicate the language that the HTQ was administered in, not the language of the paper version of the HTQ (e.g., the English paper version of the HTQ was used, but was administered orally in Urdu through the use of an interpreter, code "Urdu"). Code each language that the instrument was administered in. If there were more than five languages, code "Various".

14) Original or adapted HTQ: Code whether the HTQ was used as is, or translated and/or culturally adapted in any way. If there was no mention of this, but it was obvious that there would have been an adaptation (i.e., different number of items, or they administered the HTQ to individuals who didn't speak the language of one of the original versions) select "adaptation".

15) Was the sample of the study a sample for which the HTQ was originally validated? Original samples include: Vietnamese, Cambodian (Khmer), Lao, Arabic (Iraqi) and Russian-speaking refugees, Bosnian, Kurdish (Iraqi), Peruvian or Spanish (Peruvian) civilian survivors of war, Croatian combat soldiers, and Japanese earthquake survivors.

16) Would an adaptation have been required for the sample? In some cases, an adaptation is not explicitly described but it is obvious that an adaptation would have been needed. For example, the HTQ was administered to a Swedish-speaking sample, but no Swedish version of the HTQ has been approved by the developers. Furthermore, in some instances a cultural adaptation (or at least a cultural verification with a pre-test) would have been needed even if the language of the original HTQ was spoken by the population. For example, a sample of Liberian English speaking political prisoners completed an English version of the HTQ.

17) Type of adaptation procedure described: Only code what was *reported* by the authors.

18) Authors stated they followed the guidelines recommended by the developers of the HTQ? Only code what was *reported* by the authors.

19) Authors stated they followed guidelines developed by experts in translation/adaptation of instruments? Only code what was *reported* by the authors.

20) Translation procedure used: Only select one of the following: (a) oral translation (i.e., the HTQ was orally translated by the administrator or interpreters during the interviews); (b) forward translation (code this when they simply say "was translated" or described as a one-way, simple translation), (c) back translation (when translation was then translated back into source language), or (d) blind back translation (the back translation was done by someone who did not see the source document).

21) Did the authors report that a committee approach used during adaptation? Code "yes" if there is mention of a consensus done among a group of people regarding the final version of the adaptation (2 or more people).

22) Did the authors report that there was a cultural adaptation of the HTQ? Code "yes" if any there is any evidence of this. Code "no" if they didn't report anything about verifying cultural differences or culturally adapting.

23) Did the authors report that the adapted HTQ was pre-tested? Code "yes" if there is any evidence of a pre-test of the translation and/or cultural adaptation. Code "no" otherwise.

24) Administration of the HTQ was done by___: Indicate whether the administration of the HTQ was done by: (a) self-report; (b) nonclinicians (e.g. lay interviewers, teachers); or (c) clinicians (including research assistants/students in clinical psychology, psychiatry residents). If the administration procedures are not clearly specified, code "N/R".

Cronbach's alpha values

25-31) Cronbach's alpha values: Note the exact value reported by authors. Code "N/R" if the value was not reported.

Construct validity coefficients

32) Version of HTQ symptom scale used: Select the version of the HTQ that was used to calculate the validity coefficients.

33) Correlate: Select the construct that best described the variable being measured

34) Measure/Variable: Indicate which measure or variable the HTQ was compared to in the bivariate analysis

35) Format & values: Provide the relevant values needed to compute a correlation.

36) Other validity coefficients: If there are more than one validity coefficient, please input information in this section according to the various subheadings.

Appendix B

Reliability and Validity Generalizations Coding Sheet

Reliability and Validity Generalizations Coding Sheet

<table>
<tr><td colspan="3">Identification information</td></tr>
<tr><td colspan="2">1. Document identifier:</td><td>2. Sample of study:</td></tr>
<tr><td colspan="3">Sample and sampling characteristics</td></tr>
<tr><td>3. Sample size:</td><td>4. Age (mean):</td><td>5. Gender (% males):</td></tr>
<tr><td colspan="2">6. Culture/ethnicity of sample:</td><td>7. Country of origin (COO) of sample:</td></tr>
<tr><td colspan="2">8. Cultural orientation of COO ( ):
___ Individualist
___ Collectivist
___ Information unavailable</td><td>9. Country of study (COS):</td></tr>
<tr><td colspan="2">10. Recruitment setting ( ):
___ mental health services
___ medical treatment
___ community
___ mixed
___ not reported (N/R)</td><td>11. Trauma category ( ):
___ war, combat or terrorism
___ torture
___ human trafficking
___ sexual assault
___ nonsexual assault
___ personal illness or injury
___ witnessing death or injury
___ death, injury or illness of loved one
___ man-made disaster or fire
___ natural disaster
___ various
___ other (specify:_____)
___ none
___ N/R</td></tr>
<tr><td colspan="3">12. Forcibly displaced persons (FDP) ( )?
___ yes
___ no
___ some
___ N/R</td></tr>
</table>

Description of the HTQ administration and adaptation procedures	
13. Language(s) of administration of HTQ:	14. Original or adapted HTQ ( )? ___ original ___ adaptation ___ N/R
15. Was the sample of the study a sample for which the HTQ was originally validated? ( ) ___ yes ___ no	16. Would an adaptation/cultural verification have been required for the sample? ( ) ___ yes ___ no
17. Type of adaptation procedure described ( ): ___ translation only ___ cultural adaptation only ___ translation & cultural adaptation ___ N/R ___ not applicable (N/A) – original HTQ	18. Authors stated they followed the guidelines recommended by the developers of the HTQ? ( ) ___ yes ___ no ___ N/A (original)
19. Authors stated they followed guidelines developed by experts in translation/adaptation of instruments? ( ) ___ yes ___ no ___ N/A (original)	20. Translation procedure used ( ): ___ oral translation ___ forward translation ___ back translation ___ blind back translation ___ N/R ___ N/A (original)
21. Did the authors report that a committee approach was used during adaptation? ( ) ___ yes ___ no ___ N/A (original)	22. Did the authors report that there was a cultural adaptation of the HTQ? ( ): ___ yes ___ no ___ N/A (original)
23. Did the authors report that the adapted HTQ was pre-tested? ( ) ___ yes ___ no ___ N/A (original)	24. Administration of the HTQ was done by ( ): ___ self-report ___ clinician ___ non-clinician ___ N/R

<table>
<tr><td colspan="3">Cronbach's alpha values of the HTQ symptom scales</td></tr>
<tr><td>25. HTQ-30 α:</td><td>26. HTQ-16 α:</td><td>27. HTQ-14 cultural symptoms α:</td></tr>
<tr><td>28. Intrusion/Re-experiencing subscale (4 items) α:</td><td>29. Avoidance/numbing subscale (7 items) α:</td><td>30. Hypervigilance/arousal subscale (5 items) α:</td></tr>
<tr><td colspan="3">31. Other scale α (specify the scale):</td></tr>
<tr><td colspan="3">Construct validity coefficients</td></tr>
<tr><td colspan="3">32. Version of HTQ symptom scale used ( ):

___ 16 (PTSD – original HTQ)
___ 30 (incl. 16 PTSD – original HTQ)</td></tr>
<tr><td colspan="2">33. Correlate ( ):

Other PTSD measure___
Cumulative trauma___
Cultural symptoms___
Dissociation___
Anxiety & Depression___
Anxiety___
Depression___
Somatization___
Other:
_____</td><td>34. Measure/Variable (e.g.: HSCL-25 Anxiety, BDI-II, witnessing violence)</td></tr>
<tr><td colspan="3">35. Format & values

Correlation (r or rho): _____ Sample size: _____ p-value: _____

Kappa (k): _____ Sample size: _____

ANOVA (comparison of two independent groups) F-value: _____ Group #1 n: _____ Group #2 n: _____

Odds ratio: _____ Confidence limits: _____</td></tr>
</table>

T-test (t-value): _____ Sample size: _____

Cohen's d: _____ Sample size: _____

Other format & values:

36. Other validity coefficients: If there is more than one validity coefficient, please input information here :

Version of HTQ	Correlate	Measure/variable	Format Values

Appendix C

Studies Included in the Reliability Generalization of the HTQ (Study 1)

Studies Included in the Reliability Generalization of the HTQ (Study 1)

Ahern, J., Galea, S., Fernandez, W. G., Koci, B., Waldman, R., & Vlahov, D. (2004). Gender, social support, and posttraumatic stress in postwar Kosovo. *Journal of Nervous and Mental Disease, 192*, 762-770. doi: 10.1097/01.nmd.0000144695.02982.41

Ahmad, A., von Knorring, A. L., & Sundelin-Wahlsten, V. (2008). Traumatic experiences and post-traumatic stress disorder in Kurdistanian children and their parents in homeland and exile: An epidemiological approach. *Nordic Journal of Psychiatry, 62*, 457-463. doi: 10.1080/08039480801984305

Andersen, T. E., Andersen, L.-A. C., & Andersen, P. G. (2014). Chronic pain patients with possible co-morbid posttraumatic stress disorder admitted to multidisciplinary pain rehabilitation: A 1-year cohort study. *European Journal of Psychotraumatology, 5*. doi: 10.3402/ejpt.v5.23235

Andersen, T. E., Elklit, A., & Brink, O. (2013). PTSD symptoms mediate the effect of attachment on pain and somatization after whiplash injury. *Clinical Practice and Epidemiology in Mental Health, 9*, 75-83. doi: 10.2174/1745017901309010075

Anwar, J., Mpofu, E., Matthews, L. R., & Brock, K. E. (2013). Risk factors of posttraumatic stress disorder after an earthquake disaster. *Journal of Nervous and Mental Disease, 201*, 1045-1052. doi: 10.1097/NMD.0000000000000060

Badri, A., Crutzen, R., & Van den Borne, H. W. (2012). Exposures to war-related traumatic events and post-traumatic stress disorder symptoms among displaced Darfuri female university students: An exploratory study. *BioMed Central Public Health, 12*. doi: 10.1186/1471-2458-12-603

Bentley, J. A. (2011). *Cross-cultural assessment of psychological symptoms among Somali refugees* (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.

Bödvarsdóttir, I., & Elklit, A. (2004). Psychological reactions in Icelandic earthquake survivors. *Scandinavian Journal of Psychology, 45,* 3-13. doi: 10.1111/j.1467-9450.2004.00373.x

Bödvarsdóttir, I., & Elklit, A. (2007). Victimization and PTSD-like states in an Icelandic youth probability sample. *BMC Psychiatry, 7,* doi: 10.1186/1471-244x-7-51

Cardozo, B. L., Holtz, T .H., Kaiser, R., Gotway, C. A., Ghitis, F., Toomey, E., & Salama, P. (2005). The mental health of expatriate and Kosovar Albanian humanitarian aid workers. *Disasters, 29,* 152-170. doi: 10.1111/j.0361-3666.2005.00278.x

Carlsson, J. M., Mortensen, E. L., & Kastrup, M. (2006). Predictors of mental health and quality of life in male tortured refugees. *Nordic Journal of Psychiatry, 60,* 51-57. doi: 10.1080/08039480500504982

Christiansen, D., Bak, R., & Elklit, A. (2012). Secondary victims of rape. *Violence and Victims, 27*, 246-262. doi: 10.1891/0886-6708.27.2.246

Christiansen, D., Elklit, A., & Olff, M. (2013). Parents bereaved by infant death: PTSD symptoms up to 18 years after the loss. *General Hospital Psychiatry, 35*, 605-611. doi: 10.1016/j.genhosppsych.2013.06.006

Cleveland, J., & Rousseau, C. (2013). Psychiatric symptoms associated with brief detention of adult asylum seekers in Canada. *Canadian Journal of Psychiatry, 58,* 409-416. doi: 10.1177/070674371305800706

Correa-Velez, I., Barnett, A. G., Gifford, S. M., & Sackey, D. (2011). Health status and use of health services among recently arrived men with refugee backgrounds: a comparative

analysis of urban and regional settlement in South-east Queensland. *Australian Journal of Primary Health, 17*, 66-71. doi: 10.1071/py10051

Crescenzi, A., Ketzer, E., Van Ommeren, M., Phuntsok, K., Komproe, I., & De Jong, J. T. V. M. (2002). Effect of political imprisonment and trauma history on recent Tibetan refugees in India. *Journal of Traumatic Stress, 15*, 369-375. doi: 10.1023/A:1020129107279

Culhane, M. A. (2009). *Suffering strangers: The social construction of the torture survivor in the American healing community* (Unpublished doctoral dissertation). Boston University, Boston, MA.

de Fouchier, C., Blanchet, A., Hopkins, W., Bui, E., Ait-Aoudia, M., & Jehel, L. (2012). Validation of a French adaptation of the Harvard Trauma Questionnaire among torture survivors from sub-Saharan African countries. *European Journal of Psychotraumatology, 3*. doi: 10.3402/ejpt.v3i0.19225

Domanskaité-Gota, V., Elklit, A., & Christiansen, D. M. (2009). Victimization and PTSD in a Lithuanian national youth probability sample. *Nordic Psychology, 61*, 66-81. doi: 10.1027/1901-2276.61.3.66

Doran, J. M., Kalayjian, A., Toussaint, L., & DeMucci, J. (2012). The relationship between trauma and forgiveness in post-conflict Sierra Leone. *Psychological Trauma-Theory, Research, Practice, and Policy, 4*, 614-623. doi: 10.1037/a0025470

Eisen, E. M. (2016). *The impact on post-migration factors on PTSD and depressive symptoms among asylum seekers in the United States* (Unpublished doctoral dissertation). George Washington University, Washington, DC.

Elklit, A. (2002a). Acute stress disorder in victims of robbery and victims of assault. *Journal of Interpersonal Violence, 17*, 872-887. doi: 10.1177/0886260502017008005

Elklit, A. (2002b). Victimization and PTSD in a Danish national youth probability sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 174-181. doi: 10.1097/00004583-200202000-00011

Elklit, A., & Blum, A. (2011). Psychological adjustment one year after the diagnosis of breast cancer: A prototype study of delayed post-traumatic stress disorder. *British Journal of Clinical Psychology, 50*, 350-363. doi: 10.1348/014466510x527676

Elklit, A., & Brink, O. (2003). Acute stress disorder in physical assault victims visiting a Danish emergency ward. *Violence and Victims, 18*, 461-472. doi: 10.1891/vivi.2003.18.4.461

Elklit, A., & Kurdahl, S. (2013). The psychological reactions after witnessing a killing in public in a Danish high school. *European Journal of Psychotraumatology, 4*. doi: 10.3402/ejpt.v4i0.19826

Elklit, A., & O'Connor, M. (2005). Post-traumatic stress disorder in a Danish population of elderly bereaved. *Scandinavian Journal of Psychology, 46*, 439-445. doi: 10.1111/j.1467-9450.2005.00475.x

Elklit, A., Christiansen, D. M., Palic, S., Karsberg, S., & Eriksen, S. B. (2014). Impact of traumatic events on posttraumatic stress disorder among Danish survivors of sexual abuse in childhood. *Journal of Child Sexual Abuse, 23*, 918-934. doi: 10.1080/10538712.2014.964440

Elklit, A., Hartvig, T., & Christiansen, M. (2007). Psychological sequelae in parents of extreme low and very low birth weight infants. *Journal of Clinical Psychology in Medical Settings, 14*, 238-247. doi: 10.1007/s10880-007-9077-4

Ellis, B. H., Abdi, S. M., Lazarevic, V., White, M. T., Lincoln, A. K., Stern, J. E., & Horgan, J. G. (2016). Relation of psychosocial factors to diverse behaviors and attitudes among

Somali refugees. *American Journal of Orthopsychiatry, 86,* 393-408. doi: 10.1037/ort0000121

Eng, S., Mulsow, M., Cleveland, H., & Hart, S. L. (2009). Academic achievement among adolescents in Cambodia: Does caregiver trauma matter? *Journal of Community Psychology, 37,* 754-768. doi: 10.1002/jcop.20329

Farhood, L. F., & Dimassi, H. (2012). Prevalence and predictors for post-traumatic stress disorder, depression and general health in a population from six villages in South Lebanon. *Social Psychiatry and Psychiatric Epidemiology, 47,* 639-649. doi: 10.1007/s00127-011-0368-6

Finklestein, M., & Solomon, Z. (2009). Cumulative trauma, PTSD and dissociation among Ethiopian refugees in Israel. *Journal of Trauma and Dissociation, 10,* 38-56. doi: 10.1080/15299730802485151

Gordon, J. S., Staples, J. K., Blyta, A., Bytyqi, M., & Wilson, A. T. (2008). Treatment of posttraumatic stress disorder in postwar Kosovar adolescents using mind-body skills groups: A randomized controlled trial. *Journal of Clinical Psychiatry, 69*(9), 1469-1476.

Gordon, J. S., Staples, J. K., He, D. Y., Atti, J. A. A. (2016). Mind–body skills groups for posttraumatic stress disorder in Palestinian adults in Gaza. *Traumatology, 22,* 155-164. doi: 10.1037/trm0000081

Gupta, J., Falb, K. L., Carliner, H., Hossain, M., Kpebo, D., & Annan, J. (2014). Associations between exposure to intimate partner violence, armed conflict, and probable PTSD among women in rural Côte d'Ivoire. *PLOS One, 9,* e96300. doi: 10.1371/journal.pone.0096300

Halvorsen, J. O., & Kagee, A. (2010). Predictors of psychological sequelae of torture among South African former political prisoners. *Journal of Interpersonal Violence, 25*, 989-1005. doi: 10.1177/0886260509340547

Hansen, N. B., Eriksen, S. B., & Elklit, A. (2014). Effects of an intervention program for female victims of intimate partner violence on psychological symptoms and perceived social support. *European Journal of Psychotraumatology, 5*. doi: 10.3402/ejpt.v5.24797

Harper, G. W., Wade, R. M., Onyango, D. P., Abuor, P. A., Bauermeister, J. A., Odero, W. W., & Bailey, R. C. (2015). Resilience among gay/bisexual young men in Western Kenya: Psychosocial and sexual health outcomes. *AIDS, 29*, S261-S269. doi: 10.1097/QAD.0000000000000905

Hensen, M., Andersen, T. E., Armour, C., Elklit, A., Palic, S., & Mackrill, T. (2010). PTSD-8: A short PTSD inventory. *Clinical Practice and Epidemiology in Mental Health*, *6*, 101-108. doi: 10.2174/1745017901006010101

Hossain, M., Zimmerman, C., Abas, M., Light, M., & Watts, C. (2010). The relationship of trauma to mental disorders among trafficked and sexually exploited girls and women. *American Journal of Public Health, 100*, 2442-2449. doi: 10.2105/AJPH.2009.173229

Huijts, I., Kleijn, W. C., van Emmerik, A. A. P., Noordhof, A., & Smith, A. J. M. (2012). Dealing with man-made trauma: The relationship between coping style, posttraumatic stress, and quality of life in resettled, traumatized refugees in the Netherlands. *Journal of Traumatic Stress, 25*, 71-78. doi: 10.1002/jts.21665

Hussain, D., & Bhushan, B. (2009). Development and validation of the Refugee Trauma Experience Inventory. *Psychological Trauma: Theory, Research, Practice and Policy, 1*, 107-117. doi: 10.1037/a0016120

Ichikawa, M., Nakahara, S., & Wakai, S. (2006). Effect of post-migration detention on mental health among Afghan asylum seekers in Japan. *Australian and New Zealand Journal of Psychiatry, 40*, 341-346. doi: 10.1080/j.1440-1614.2006.01800.x

Jakobsen, M., Thoresen, S., & Johansen, L. E. E. (2011). The validity of screening for post-traumatic stress disorder and other mental health problems among asylum seekers from different countries. *Journal of Refugee Studies, 24*, 171-186. doi: 10.1093/jrs/feq053

Jind, L., Elklit, A., & Christiansen, D. (2010). Cognitive schemata and processing among parents bereaved by infant death. *Journal of Clinical Psychology in Medical Settings, 17*, 366-377. doi: 10.1007/s10880-010-9216-1

Jones, L., & Kafetsios, K. (2002). Assessing adolescent mental health in war-affected societies: The significance of symptoms. *Child Abuse & Neglect, 26*, 1059-1080. doi: 10.1016/S0145-2134(02)00381-2

Kagee, K. (2005). Symptoms of distress and posttraumatic stress among South African former political detainees. *Ethnicity & Health, 10*, 169-179. doi: 10.1080/13557850500071244

Kanninen, K., Punamaki, S. L., & Qouta, S. (2002). The relation of appraisal, coping efforts, and acuteness of trauma to PTS symptoms among former political prisoners. *Journal of Traumatic Stress, 15*, 245-253. doi: 10.1023/A:1015211529584

Kanninen, K., Salo, J., & Punamaki, R. L. (2000). Attachment patterns and working alliance in trauma therapy for victims of political violence. *Psychotherapy Research, 10*, 435-449. doi: 10.1093/ptr/10.4.435

Karsberg, S., Armour, C., & Elklit, A. (2014). Patterns of victimization, suicide attempt, and posttraumatic stress disorder in Greenlandic adolescents: A latent class analysis. *Social*

Psychiatry and Psychiatric Epidemiology, 49, 1389-1399. doi: 10.1007/s00127-014-0890-4

Keller, A., Lhewa, D., Rosenfeld, B., Sachs, E., Aladjem, A., Cohen, I., . . . Portetfield, K. (2006). Traumatic experiences and psychological distress in an urban refugee population seeking treatment services. *Journal of Nervous and Mental Disease, 194*, 188-194. doi: 10.1097/01.nmd.0000202494.75723.83

Kleijn, W. C., Hovens, J. E., & Rodenburg, J .J. (2001). Posttraumatic stress symptoms in refugees: Assessments with the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist-25 in different languages. *Psychological Reports, 88*, 527-532. doi: 10.2466/PR0.88.2.527-532

Kristensen, T. E., Elklit, A., Karstoff, K.-A., & Palic, S. (2014). Predicting chronic posttraumatic stress disorder in bereaved relatives: A 6-month follow-up study. *American Journal of Hospice and Palliative Medicine, 31*, 396-405. doi: 10.1177/1049909113490066

Kuo, C., Reddy, M. K., Operario, D., Cluver, L., & Stein, D. J. (2013). Posttraumatic stress symptoms among adults caring for orphaned children in HIV-endemic South Africa. *AIDS and Behavior, 17*, 1755-1763. doi: 10.1007/s10461-013-0461-9

Leaman, S. C. (2009). *Risk factors for psychological distress and uses of religious coping* (Unpublished doctoral dissertation). The George Washington University, Washington, DC.

Lhewa, D. W. (2010). C*oping and distress among Tibetan survivors of torture and refugee trauma* (Unpublished doctoral dissertation). Boston University, Boston, MA.

Machisa, M. T., Christofides, N., & Jewkes, R. (2016). Structural pathways between child abuse, poor mental health outcomes and male-perpetrated intimate partner violence (IPV). *PLoS ONE, 11*, e0150986. doi:10.1371/journal.pone.0150986

Mai, L. G. (2010). *Exploration of the intergenerational transmission of historical trauma in Vietnamese Americans* (Unpublished doctoral dissertation). California Institute of Integral Studies, San Francisco, CA.

Michalopoulos, L. (2013). *Exploration of cross-cultural adaptability of PTSD among trauma survivors in Northern Iraq, Thailand, and Democratic Republic of Congo: Application of item response theory and classical test theory* (Unpublished doctoral dissertation). University of Maryland, Baltimore, MD.

Mollica, R. F., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., & Lavelle, J. (1992). The Harvard Trauma Questionnaire: Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *Journal of Nervous and Mental Disease, 180*, 111-116. doi: 10.1097/00005053-199202000-00008

Mollica, R. F., Chernoff, M. C., Berthold, S. M., Lavelle, J., Lyoo, I. K., & Renshaw, P. (2014). The mental health sequelae of traumatic head injury in South Vietnamese ex-political detainees who survived torture. *Comprehensive Psychiatry, 55*, 1626-1638. doi: 10.1016/j.comppsych.2014.04.014

Mollica, R. F., McInnes, K., Sarajlic, N., Lavelle, J., Sarajlic, J., & Massagli, M. P. (1999). Disability associated with psychiatric comorbidity and health status in Bosnian refugees living in Croatia. *Journal of the American Medical Association, 282*, 433-439. doi: 10.1001/jama.282.5.433

Morina, N., & Ford, J. D. (2008). Complex sequelae of psychological trauma among Kosovar civilian war victims. *International Journal of Social Psychiatry, 54*, 425-436. doi: 10.1177/0020764008090505

Nickerson, A., Bryant, R. A., Brooks, R., Steel, Z., Silove, D., & Chen, J. (2011). The familial influence of loss and trauma on refugee mental health: A multilevel path analysis. *Journal of Traumatic Stress, 24*, 25-33. doi: 10.1002/jts.20608

O'Connor, M. (2010). PTSD in older bereaved people. *Aging and Mental Health, 14*, 670-678. doi: 10.1080/13607860903311725

O'Connor, M., & Elklit, A. (2008). Attachment styles, traumatic events, and PTSD: a cross-sectional investigation of adult attachment and trauma. *Attachment & Human Development, 10*, 59-71. doi: 10.1080/14616730701868597

Opaas, M., & Hartmann, E. (2013). Rorschach assessment of traumatized refugees: An exploratory factor analysis. *Journal of Personality Assessment, 95*, 457-470. doi: 10.1080/00223891.2013.781030

Oruc, L., Kapetanovic, A., Pojskic, N., Miley, K., Forstbauer, S., Mollica, R. F., & Henderson, D. C. (2008). Screening for PTSD and depression in Bosnia and Herzegovina: Validating the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist. *International Journal of Culture and Mental Health, 1*, 105-116. doi: 10.1080/17542860802456620

Petersen, T., Elklit, A., & Olesen, J. G. (2010). Victimization and PTSD in a Faroese youth total-population sample. *Scandinavian Journal of Psychology, 51*, 56-62. doi: 10.1111/j.1467-9450.2009.00728.x

Pham, K. T. (2013). *Trauma transmission between first-generation and second-generation Vietnamese refugees and immigrants in the United States: A study of mother and daughter* (Unpublished doctoral dissertation). Palo Alto University, Palo Alto, CA.

Renner, W., Salem, I., & Ottomeyer, K. (2006). Cross-cultural validation of measures of traumatic symptoms in groups of asylum seekers from Chechnya, Afghanistan, and West Africa. *Social Behavior and Personality, 34*, 1101-1114. doi: 10.2224/sbp.2006.34.9.1101

Roberts, B., Damundu, E. Y., Lomoro, O., & Sondorp, E. (2009). Post-conflict mental health needs: A cross-sectional survey of trauma, depression and associated factors in Juba, Southern Sudan. *BioMed Central Psychiatry, 9,* 7. doi: 10.1186/1471-244X-9-7

Roberts, B., Ocaka, K. F., Browne, J., Oyok, T., & Sondorp, E. (2008). Factors associated with post-traumatic stress disorder and depression amongst internally displaced persons in northern Uganda. *BMC Psychiatry, 8,* 38. doi: 10.1186/1471-244x-8-38

Roth, G., Ekbad, S., & Agren, H. (2006). A longitudinal study of PTSD in a sample of adult mass-evacuated Kosovars, some of whom returned to their home country. *European Psychiatry, 21*, 152-159. doi: 10.1016/j.eurpsy.2005.11.004

Rothon, C., Stansfeld, S. A., Mathews, C., Kleinhans, A., Clark, C., Lund, C., & Flisher, A. J. (2011). Reliability of self-report questionnaires for epidemiological investigations of adolescent mental health in Cape Town, South Africa. *Journal of Child and Adolescent Mental Health, 23*, 119-128. doi: 10.2989/17280583.2011.634551

Sachs, E., Rosenfeld, B., Lhewa, D., Rasmussen, A., & Keller, A. (2008). Entering exile: Trauma, mental health, and coping among Tibetan refugees arriving in Dharamsala, India. *Journal of Traumatic Stress, 21*, 199-208. doi: 10.1002/jts.20324

Schubert, S. J., Lee, C. W., de Araujo, G., Butler, S. R., Taylor, G., & Drummond, P. D. (2016). The effectiveness of Eye Movement Desensitization and Reprocessing Therapy to treat symptoms following trauma in Timor Leste. *Journal of Traumatic Stress, 29,* 141-148. doi: 10.1002/jts.22084

Semb, O., Henningsson, M., Fransson, P., & Sundbom, E. (2009). Trauma-related symptoms after violent crime: The role of risk factors before, during and eight months after victimization. *The Open Psychology Journal, 2*, 77-88. doi: 10.2174/1874350100902010077

Silove, D., Brooks, R., Steel Bateman, C., Steel, Z., Fonseca, Z., Amaral, C.,… Soosay, I. (2010). Social and trauma-related pathways leading to psychological distress and functional limitations four years after the humanitarian emergency in Timor Leste. *Journal of Traumatic Stress, 23,* 151-160. doi: 10.1002/jts.20499

Silove, D., Rees, S., Tay, A. K., da Costa, Z. M., Soares Savio, E., Soares, C., & Tol, W. (2015). Pathways to perinatal depressive symptoms after mass conflict in Timor-Leste: A modelling analysis using cross-sectional data. *Lancet Psychiatry, 2*, 161-167. doi: 10.1016/S2215-0366(14)00054-6

Slewa-Younan, S., Chippendale, K., Heriseanu, A., Lujic, S., Atto, J., & Raphael, B. (2012). Measures of psychophysiological arousal among resettled traumatized Iraqi refugees seeking psychological treatment. *Journal of Traumatic Stress, 25,* 348-352. doi: 10.1002/jts.21694

Slewa-Younan, S., Mond, J. M., Bussion, E., Melkonian, M., Mohammad, Y., Dover, H., … Jorm, A. F. (2015). Psychological trauma and help seeking behaviour amongst resettled

Iraqi refugees in attending English tuition classes in Australia. *International Journal of Mental Health Systems, 9*. doi: 10.1186/1752-4458-9-5.

Sondergaard, H. P., Ekblad, S., & Theorell, T. (2001). Self-reported life event patterns and their relation to health among recently resettled Iraqi and Kurdish refugees in Sweden. *Journal of Nervous and Mental Disease, 189*(12), 838-845.

Southivong, B., Ichikawa, M., Nakahara, S., & Southivong, C. (2013). A cross-sectional community study of posttraumatic stress disorder and social support in Lao People's Democratic Republic. *Bulletin of the World Health Organization, 91*, 765-772. doi: 10.2471/BLT.12.115311

Soysa, C. K. (2013). War and tsunami PTSD responses in Sri Lankan children: Primacy of re-experiencing and arousal compared to avoidance-numbing. *Journal of Aggression, Maltreatment and Trauma, 22*, 896-915. doi: 10.1080/10926771.2013.824056

Suarez, E. B. (2011). *Surviving the "Sasachacuy Tiempu "[difficult times]: The resilience of Quechua women in the aftermath of the Peruvian armed conflict* (Unpublished doctoral dissertation). University of Toronto, Toronto, ON.

Veer, I. M., Oei, N. Y. L., van Buchem, M. A., Spinhoven, P., Elzinga, B. M., & Rombouts, S. A. R. B. (2015). Evidence for smaller right amygdala volumes in posttraumatic stress disorder following childhood trauma. *Psychiatry Research: Neuroimaging, 233*, 436-442. doi: 10.1016/j.pscychresns.2015.07.016

Vervliet, M., Meyer Demott, M. A., Jakobsen, M., Broekaert, E., Heir, T., & Derluyn, I. (2014). The mental health of unaccompanied refugee minors on arrival in the host country. *Scandinavian Journal of Psychology, 33*, 33-37. doi: 10.1111/sjop.12094

Vloeberghs, E., van der Kwaak, A., Knipscheer, J., & van den Muijsenbergh (2012). Coping and chronic psychosocial consequences of female genital mutilation in the Netherlands. *Ethnicity and Health, 17*, 677-695. doi: 10.1080/13557858.2013.771148

Vromans, L., Schweitzer, R. D., & Brough, M. (2012). The Multidimensional Loss Scale: Validating a cross-cultural instrument for measuring loss. *Journal of Nervous and Mental Disease, 200,* 349-357. doi: 10.1097/NMD.0b013e31824cc458

Wagner, J., Burke, G., Kuoch, T., Scully, M., Armeli, S., & Rajan, T. V. (2013). Trauma, healthcare access, and health outcomes among Southeast Asian refugees in Conneticut. *Journal of Immigrant and Minority Health, 15*, 1065-1072. doi: 10.1007/s10903-012-9715-2

Ward, C. L., Flisher, A. J., Chrisostomos, Z., Muller, M., & Lombard, C. J. (2004). Measuring adolescents' exposure to violence and related PTSD symptoms: Reliability of an adaptation of the Harvard Trauma Questionnaire. *Journal of Child and Adolescent Mental Health, 16*, 31-37. doi; 10.2989/17280580409486561

Welsh, E. A. (2012). *Predictors of mental health and treatment engagement in torture survivors* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.

Yeomans, P. D. (2008). *The effect of PTSD psychoeducation on the nature and severity of traumatic stress symptoms in a Burundian sample* (Unpublished doctoral dissertation). Drexel University, Philadelphia, PA.

Appendix D

Studies Included in the Validity Generalization of the HTQ (Study 2)

Studies Included in the Validity Generalization of the HTQ (Study 2)

Ahmad, A., Sofi, M.A., Sundelin-Wahlsten, V. & von Knorring, A.-L. (2000). Posttraumatic stress disorder in children after the military operation "Anfal" in Iraqi Kurdistan. *European Child & Adolescent Psychiatry, 9*, 235-243. doi: 10.1007/s007870070026

Aragona, M., Pucci, D., Mazzetti, M., Maisano, B., & Geraci, S. (2013). Traumatic events, post-migration living difficulties and post-traumatic symptoms in first generation immigrants: A primary care study. *Annali dell'Istituto Superiore di Sanità, 49*, 169-175. doi: 10.4415/ANN_13_02_08

Bentley, J. A. (2011). *Cross-cultural assessment of psychological symptoms among Somali refugees* (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.

Buhmann, C. B. (2014). Traumatized refugees: Morbidity, treatment and predictors of outcome. *Danish Medical Journal, 61*(8), B4871.

Carswell, K., Blackburn, P., & Barker, C. (2011). Relationship between trauma, post-migration problems and the psychological well-being of refugees and asylum seekers. *International Journal of Social Psychiatry, 57*, 107-119. doi: 10.1177/0020764008105699

Chu, T., Keller, A. S., & Rasmussen, A. (2013). Effects of post-migration factors on PTSD outcomes among immigrant survivors of political violence. *Journal of Immigrant and Minority Health, 15*, 890-897. doi: 10.1007/s10903-012-9696-1

Culhane, M. A. (2009). *Suffering strangers: The social construction of the torture survivor in the American healing community* (Unpublished doctoral dissertation). Boston University, Boston, MA.

Dewaraja, R., & Kawamura, N. (2006). Trauma intensity and posttraumatic stress: Implications of the tsunami experience in Sri Lanka for the management of future disasters. *International Congress Series, 1287*, 69-73. doi: 10.1016/j.ics.2005.11.098

Domanskaité-Gota, V., Elklit, A., & Christiansen, D. M. (2009). Victimization and PTSD in a Lithuanian national youth probability sample. *Nordic Psychology, 61*, 66-81. doi: 10.1027/1901-2276.61.3.66

Doran, J. M., Kalayjian, A., Toussaint, L., & DeMucci, J. (2012). The relationship between trauma and forgiveness in post-conflict Sierra Leone. *Psychological Trauma-Theory Research Practice and Policy*, *4*, 614-623. doi: 10.1037/a0025470

Dworkin, J., Prescott, M., Jamal, R., Hardawan, S. A., Abdullah, A., & Galea, S. (2008). The long-term psychosocial impact of a surprise chemical weapons attack on civilians in Halabja, Iraqi Kurdistan. *Journal of Nervous and Mental Disease, 196*, 772-775. doi: 10.1097/NMD.0b013e3181878b69

Ekblad, S., Prochazka, H. & Roth, G. (2002). Psychological impact of torture: A 3 month follow-up of mass-evacuated Kosovan adults in Sweden: Lessons learnt for prevention. *Acta Psychiatrica Scandinavica, 106*, 30-36. doi: 10.1034/j.1600- 0447.106.s412.7.x

Elklit, A., Christiansen, D. M., Palic, S., Karsberg, S., & Eriksen, S. B. (2014). Impact of traumatic events on posttraumatic stress disorder among Danish survivors of sexual abuse in childhood. *Journal of Child Sexual Abuse, 23*, 918-934. doi: 10.1080/10538712.2014.964440

Elklit, A., & Kurdahl, S. (2013). The psychological reactions after witnessing a killing in public in a Danish high school. *European Journal of Psychotraumatology, 4*. doi: 10.3402/ejpt.v4i0.19826

Elklit, A., Nielsen, L. H., Lasgaard, M., & Duch, C. (2013). A cartoon-based measure of PTSD symptomatology in children exposed to a disaster. *Journal of Loss & Trauma, 18*, 54-63. doi: 10.1080/15325024.2012.679125

Eng, S., Mulsow, M., Cleveland, H., & Hart, S. L. (2009). Academic achievement among adolescents in Cambodia: Does caregiver trauma matter? *Journal of Community Psychology, 37*, 754-768. doi: 10.1002/jcop.20329

Farhood, L. F., & Dimassi, H. (2012). Prevalence and predictors for post-traumatic stress disorder, depression and general health in a population from six villages in South Lebanon. *Social Psychiatry and Psychiatric Epidemiology, 47*, 639-649. doi: 10.1007/s00127-011-0368-6

Farhood, L., Dimassi, H., & Lehtinen, T. (2006). Exposure to war-related traumatic events, prevalence of PTSD, and general psychiatric morbidity in a civilian population from southern Lebanon. *Journal of Transcultural Nursing, 17*, 333-340. doi: 10.1177/1043659606291549

Fox, S. H., & Willis, M. S. (2009). Initiatory mental health assessments for Dinka and Nuer refugees from Sudan. *Journal of Immigrant and Refugee Studies, 7*, 159-179. doi: 10.1080/15562940902935621

Galea, S., Rockers, P. C., Saydee, G., Macauley, R., Varpilah, S. T., & Kruk, M. E. (2010). Persistent psychopathology in the wake of civil war: Long-term posttraumatic stress disorder in Nimba County, Liberia. *American Journal of Public Health, 100*, 1745-1751. doi: 10.2105/ajph.2009.179697

Geltman, P. L., Grant-Knight, W., Ellis, H., & Landgraf, J. M. (2008). The "Lost Boys" of Sudan: Use of health services and functional health outcomes of unaccompanied refugee

minors resettled in the U. S. *Journal of Immigrant and Minority Health, 10,* 389-396. doi: 10.1007/s10903-007-9110-6

Gordon, J. S., Staples, J. K., He, D. Y., Atti, J. A. A. (2016). Mind–body skills groups for posttraumatic stress disorder in Palestinian adults in Gaza. *Traumatology, 22*, 155-164. doi: 10.1037/trm0000081

Grant-Knight, W., Geltman, P. L., & Ellis, H. B. (2009). Physical and mental health functioning in Sudanese unaccompanied minors. In W. Grant-Knight, P. L. Geltman & H. B. Ellis (Eds.), *Treating traumatized children: Risk, resilience and recovery* (pp. 102-116). New York: Routledge.

Gupta, J., Falb, K. L., Carliner, H., Hossain, M., Kpebo, D., & Annan, J. (2014). Associations between exposure to intimate partner violence, armed conflict, and probable PTSD among women in rural Côte d'Ivoire. *PLOS One, 9*, e96300. doi: 10.1371/journal.pone.0096300

Hadley, C., Tegegn, A., Tessema, F., Cowan, J. A., Asefa, M., & Galea, S. (2008). Food insecurity, stressful life events and symptoms of anxiety and depression in east Africa: Evidence from the Gilgel Gibe growth and development study. *Journal of Epidemiology and Community Health, 62,* 980-986. doi: 10.1136/jech.2007.068460

Halvorsen, J. O., & Kagee, A. (2010). Predictors of psychological sequelae of torture among South African former political prisoners. *Journal of Interpersonal Violence, 25*, 989-1005. doi: 10.1177/0886260509340547

Hooberman, J. B., Rosenfeld, B., Lhewa, D., Rasmussen, A., & Keller, A. (2007). Classifying the torture experiences of refugees living in the United States. *Journal of Interpersonal Violence, 22*, 108-123. doi: 10.1177/0886260506294999

Hossain, M., Zimmerman, C., Abas, M., Light, M., & Watts, C. (2010). The relationship of trauma to mental disorders among trafficked and sexually exploited girls and women. *American Journal of Public Health, 100,* 2442-2449. doi: 10.2105/AJPH.2009.173229

Huijts, I., Kleijn, W. C., van Emmerik, A. A. P., Noordhof, A., & Smith, A. J. M. (2012). Dealing with man-made trauma: The relationship between coping style, posttraumatic stress, and quality of life in resettled, traumatized refugees in the Netherlands. *Journal of Traumatic Stress, 25,* 71-78. doi: 10.1002/jts.21665

Husain, F., Anderson, M., Cardozo, B. L., Becknell, K., Blanton, C., Araki, D., & Vithana, E. K. (2011). Prevalence of war-related mental health conditions and association with displacement status in postwar Jaffna District, Sri Lanka. *Journal of the American Medical Association, 306,* 522-531. doi: 10.1001/jama.2011.1052

Hussain, D., & Bhushan, B. (2009). Development and validation of the Refugee Trauma Experience Inventory. *Psychological Trauma: Theory, Research, Practice and Policy, 1,* 107-117. doi: 10.1037/a0016120

James, L. E., Noel, J. R., & Pierre, Y. M. (2014). A mixed-methods assessment of the experiences of lay mental health workers in post-earthquake Haiti. *American Journal of Orthopsychiatry, 84,* 152-163. doi: 10.1037/h0099387

Jones, L., & Kafetsios, K. (2002). Assessing adolescent mental health in war-affected societies: The significance of symptoms. *Child Abuse & Neglect, 26,* 1059-1080. doi: 10.1016/S0145-2134(02)00381-2

Joska, J. A., Fincham, D. S., Stein, D. J., Paul, R. H., & Seedat, S. (2010). Clinical correlates of HIV-associated neurocognitive disorders in South Africa. *AIDS and Behavior, 14,* 371-378. doi: 10.1007/s10461-009-9538-x

Keller, A., Lhewa, D., Rosenfeld, B., Sachs, E., Aladjem, A., Cohen, I., . . . Portetfield, K. (2006). Traumatic experiences and psychological distress in an urban refugee population seeking treatment services. *Journal of Nervous and Mental Disease, 194,* 188-194. doi: 10.1097/01.nmd.0000202494.75723.83

Kumar, M. S., Murhekar, M. V., Hutin, Y., Subramanian, T., Ramachandran, V., & Gupte, M. D. (2007). Prevalence of posttraumatic stress disorder in a coastal fishing village in Tamil Nadu, India, after the December 2004 tsunami. *American Journal of Public Health, 97*, 99-101. doi: 10.2105/ajjph.2005.071167

Kun, P., Han, S., Chen, X., & Yao, L. (2009). Prevalence and risk factors for posttraumatic stress disorder: A cross-sectional study among survivors of the Wenchuan 2008 earthquake in China. *Depression and Anxiety, 26,* 1134-1140. doi: 10.1002/da.20612

Leaman, S. C. (2009). *Risk factors for psychological distress and uses of religious coping* (Unpublished doctoral dissertation). The George Washington University, Washington, DC.

Leaman, S. C., & Gee, C. B. (2012). Religious coping and risk factors for psychological distress among African torture survivors. *Psychological Trauma: Theory, Research, Practice, and Policy, 4*, 457-465. doi: 10.1037/a0026622

Lie, B. (2002). A 3-year follow-up study of psychosocial functioning and general symptoms in settled refugees. *Acta Psychiatrica Scandinavica, 106*, 415-425. doi: 10.1034/j.1600-0447.2002.01436.x

Mai, L. G. (2010). *Exploration of the intergenerational transmission of historical trauma in Vietnamese Americans* (Unpublished doctoral dissertation). California Institute of Integral Studies, San Francisco, CA.

Matthey, S., Silove, D., Barnett, B., Fitzgerald, M.H. & Mitchell, P. (1999). Correlates of depression and PTSD in Cambodian women with young children: A pilot study. *Stress Medicine, 15,* 103-107. doi: 10.1002/(SICI)1099-1700(199904)15:2<103::AID-SMI791>3.0.CO;2-T

Meffert, S. M., Abdo, A. O., Abd Alla, O. A., Elmakki, Y. O. M., Omer, A. A., Yousif, S., … Marmar, C. R. (2014). A pilot randomized controlled trial of interpersonal psychotherapy for Sudanese refugees in Cairo, Egypt. *Psychological Trauma: Theory, Research, Practice, and Policy, 6,* 240-249. doi: 10.1037/a0023540

Mghir, R. & Raskin, A. (1999). The psychological effects of the war in Afghanistan on young Afghan refugees from different ethnic backgrounds. *International Journal of Social Psychiatry, 45*, 29-36. doi: 10.1177/002076409904500104

Möhlen, H., Parzer, P., Resch, F., & Brunner, R. (2005). Psychosocial support for war-traumatized child and adolescent refugees: Evaluation of a short-term treatment program. *Australian and New Zealand Journal of Psychiatry, 39,* 81-87. doi: 10.1080/j.1440-1614.2005.01513.x

Mollica, R. F., McInnes, K., Pham, T., Smith, M.C., Murphy, E. & Lin, L. (1998). The dose-effect relationships between torture and psychiatric symptoms in Vietnamese ex-political detainees and a comparison group. *Journal of Nervous and Mental Disease, 186*, 543-553. doi: 10.1097/00005053-199809000-00005

Mollica, R. F., McInnes, K., Poole, C. & Tor, S. (1998). Dose-effect relationships of trauma to symptoms of depression and post-traumatic stress disorder among Cambodian survivors of mass violence. *British Journal of Psychiatry, 173,* 482-488. doi: 10.1192/bjp.173.6.482

Morina, N., & Ford, J. D. (2008). Complex sequelae of psychological trauma among Kosovar civilian war victims. *International Journal of Social Psychiatry, 54*, 425-436. doi: 10.1177/0020764008090505

Nicholson, B.L. (1997). The influence of pre-emigration and post-migration stressors on mental health: A study of Southeast Asian refugees. *Social Work Research, 21*(1), 19-31.

O'Connor, M. (2010). PTSD in older bereaved people. *Aging and Mental Health, 14*, 670-678. doi: 10.1080/13607860903311725

O'Connor, M. (2010). PTSD in the older bereaved people. *Aging and Mental Health, 14*, 310-318. doi: 10.1080/13607860903228770

O'Connor, M., Nickerson, A., Aderka, I. M., & Bryant, R. A. (2015). The temporal relationship between change in symptoms of prolonged grief and posttraumatic stress following old age spousal bereavement. *Depression and Anxiety, 32*, 335-340. doi: 10.1002/da.22349

Oruc, L., Kapetanovic, A., Pojskic, N., Miley, K., Forstbauer, S., Mollica, R. F., & Henderson, D. C. (2008). Screening for PTSD and depression in Bosnia and Herzegovina: Validating the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist. *International Journal of Culture and Mental Health, 1*, 105-116. doi: 10.1080/17542860802456620

Ovuga, E., Oyok, T. O., & Moro, E. B. (2008). Posttraumatic stress disorder among former child soldiers attending a rehabilitative service and primary school education in northern Uganda. *African Health Sciences, 8*(3), 136-141.

Pham, K. T. (2013). *Trauma transmission between first-generation and second-generation Vietnamese refugees and immigrants in the United States: A study of mother and daughter* (Unpublished doctoral dissertation). Palo Alto University, Palo Alto, CA.

Phipps, R. M. (2014). *Transgenerational transmission of trauma in second generation latino children and adolescents* (Unpublished doctoral dissertation). University of Mississippi, Oxford, MS.

Prot, K. (2010). Late effects of trauma: PTSD in Holocaust survivors. *Journal of Loss and Trauma, 15,* 28-42. doi: 10.1080/15325020902925506

Roth, G., Ekbad, S., & Agren, H. (2006). A longitudinal study of PTSD in a sample of adult mass-evacuated Kosovars, some of whom returned to their home country. *European Psychiatry, 21,* 152-159. doi: 10.1016/j.eurpsy.2005.11.004

Sabin, M., Sabin, K., Kim, H. Y., Vergara, M., & Varese, L. (2006). The mental health status of Mayan refugees after repatriation to Guatemala. *Pan American Journal of Public Health, 19,* 163-171. doi: 10.1590/s1020-49892006000300004

Sachs, E. (2011). *Assessing the impact of psychological distress on the daily functioning of refugees: A high-risk symptom profile for disability among non-Western trauma survivors* (Unpublished doctoral dissertation). Fordham University, New York.

Sachs, E., Rosenfeld, B., Lhewa, D., Rasmussen, A., & Keller, A. (2008). Entering exile: Trauma, mental health, and coping among Tibetan refugees arriving in Dharamsala, India. *Journal of Traumatic Stress, 21,* 199-208. doi: 10.1002/jts.20324

Schweitzer, R. D., Brough, M., Vromans, L., Asic-Kobe, M. (2011). Mental health of newly arrived Burmese refugees in Australia: Contributions of pre-migration and post-migration experience. *Australian and New Zealand Journal of Psychiatry, 45,* 299-307. doi: 10.3109/00048674.2010.543412

Semb, O., Henningsson, M., Fransson, P., & Sundbom, E. (2009). Trauma-related symptoms after violent crime: The role of risk factors before, during and eight months after

victimization. *The Open Psychology Journal, 2*, 77-88. doi: 10.2174/1874350100902010077

Sheikh, T. L., Abdulaziz, M., Agunbiade, S., Joseph, I., Ebiti, B., & Adekeye, O. (2015). Correlates of depression among internally displaced persons after post-election violence in Kaduna, North Western Nigeria. *Journal of Affective Disorders, 170*, 46-51. doi: 10.1016/j.jad.2014.08.050

Silove, D., Brooks, R., Steel Bateman, C., Steel, Z., Fonseca, Z., Amaral, C.,… Soosay, I. (2010). Social and trauma-related pathways leading to psychological distress and functional limitations four years after the humanitarian emergency in Timor Leste. *Journal of Traumatic Stress, 23*, 151-160. doi: 10.1002/jts.20499

Silove, D., Steel, Z., McGorry, P., Miles, V. & Drobny, J. (2002). The impact of torture on post-traumatic stress symptoms in war-affected Tamil refugees and immigrants. *Comprehensive Psychiatry, 43*, 49-55. doi: 10.1053/comp.2002.29843

Slewa-Younan, S., Chippendale, K., Heriseanu, A., Lujic, S., Atto, J., & Raphael, B. (2012). Measures of psychophysiological arousal among resettled traumatized iraqi refugees seeking psychological treatment. *Journal of Traumatic Stress, 25*, 348-352. doi: 10.1002/jts.21694

Steel, Z., & Silove, D. (2000). The psychosocial cost of seeking and granting asylum. In Z. Steel & D. Silove (Eds.), *International handbook of human response to trauma* (pp. 421-438). New York: Kluwer Academic/Plenum Publishers.

Suarez, E. B. (2011). *Surviving the "Sasachacuy Tiempu "[difficult times]: The resilience of Quechua women in the aftermath of the Peruvian armed conflict* (Unpublished doctoral dissertation). University of Toronto, Toronto, ON.

Suarez, E. B. (2013). Two decades later: The resilience and posttraumatic responses of Indigenous Quechua girls and adolescents in the aftermath of the Peruvian armed conflict. *Child Abuse and Neglect, 37*, 200-210. doi: 10.1016/j.chiabu.2012.09.011

Tay, A. K., Rees, S., Chen, J., Kareth, M., & Silove, D. (2015). The structure of post-traumatic stress disorder and complex post-traumatic stress disorder amongst West Papuan refugees. *BMC Psychiatry, 15*, 111. doi: 10.1186/s12888-015-0480-3

Toar, M., O'Brien, K. K., & Fahey, T. (2009). Comparison of self-reported health and healthcare utilisation between asylum seekers and refugees: An observational study. *BioMed Central Public Health, 9*. doi: 10.1186/1471-2458-9-214

Tomazinis, M. L. (2009). *Sub-Saharan African refugees in Hong Kong: Pre- and post-migration experiences, depression, anxiety and PTSD* (Unpublished doctoral dissertation). Alliant International University, Los Angeles, CA.

van Griensven, F., Chakkraband, M. L. S., Thienkrua, W., Pengjuntr, W., Cardozo, B. L., Tantipiwatanaskul, P., . . . Tappero, J. W. (2006). Mental health problems among adults in tsunami-affected areas in southern Thailand. *Journal of the American Medical Association, 296*, 537-548. doi: 10.1001/jama.296.5.537

van Wyk, A., Schweitzer, R. D., Brough, M., Vromans, L., & Murray, K. (2012). A longitudinal study of mental health in refugees from Burma: The impact of therapeutic interventions. *Australian and New Zealand Journal of Psychiatry, 46*, 995. doi: 10.1177/0004867412443059

Vromans, L., Schweitzer, R. D., & Brough, M. (2012). The Multidimensional Loss Scale: Validating a cross-cultural instrument for measuring loss. *Journal of Nervous and Mental Disease, 200*, 349-357. doi: 10.1097/NMD.0b013e31824cc458

Welsh, E. A. (2012). *Predictors of mental health and treatment engagement in torture survivors* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.

Wyshak, G. (1994). The relation between change in reports of traumatic events and symptoms of psychiatric distress. *General Hospital Psychiatry, 16*, 290-297. doi: 10.1016/0163-8343(94)90009-4