Validation of the Spiritual Harm and Abuse Scale with the Rating Scale Model

Jack Huber, Swedish Medical Center
Dan Koch, Northwest University

8/13/2022

Table of contents

Abstract

Spiritual harm and abuse has been a serious problem that has begun to gain the attention of the clinical and psychological research community. Some research has attempted to define spiritual harm and abuse and documented lasting emotional and psychological damage to victims of spiritual abuse. A valid scale of spiritual and religious abuse is needed, and the nascent Spiritual Harm and Abuse Scale (SHAS) has potential to meet this need. Koch and Edstrom (2022) developed an item pool and conducted an initial validation study of the SHAS which established the multidimensional structure of the item pool. This study builds on Koch and Edstrom (2022) to further validate the SHAS by using the Rasch Rating Scale Model.

Introduction

Spiritual and religious abuse has been a serious problem which is gaining the attention of the psychological research community. Koch and Edstrom (2022) define spiritual abuse as "a type of emotional and psychological abuse perpetrated by a religious leader or group and/or with a religious or spiritual component, usually involving coercion or control." Such abuse can range from psychologically controlling behaviors like quoting scripture or invoking divine rationale for mistreatment (Oakley et al., 2018); to diverse emotionally abusive behaviors such as rejecting, isolating, terrorizing, ignoring, corrupting, verbally assaulting, and over-pressuring (Pearl, 1994); to sexual or other physical forms of abuse by a clergy member (Doyle, 2006). Spiritual abuse can also range in specificity from abusive acts by religious leaders to abuse perpetrated by a religious group or person representing a religious group with a religious or spiritual component (Swindle, 2017). Spiritual abuse may also include theologically sanctioned mistreatment of a group of people such LGBTQ+ persons (Foster, Bowland, & Vosler, 2015).

Spiritual and religious abuse has serious negative effects (Ward, 2011; Swindle, 2017; Oakley & Kinmond, 2014; Stevens et al. 2019). Religious leaders may be on a par with parents in their potential for abusive harm. Further, trauma that takes place in a religious or spiritual setting may deprive victims of religion and spirituality as a coping strategy. There are also concerns related to secondary trauma for clinicians working with R/S abused clients (Gubi & Jacobs, 2009).

In light of these concerns, a valid scale of spiritual harm and abuse is needed. Practitioners who treat clients who may be victims of religious and spiritual abuse need an instrument for clinical use. Clinicians could use such a tool as part of the intake process, to evaluate their client's score in light of established norms. In addition, researchers need a reliable measurement of spiritual abuse to examine relationships between spiritual abuse and any number of other psychological, health, or other measures.

Drawing from a recent unpublished scale of religious and abuse (Keller 2016) and other scales of abuse (Kira, et al. 2008; Sanders & Becker-Lausen, 1995; Carlson, et al. 2011; Briere & Runtz, 1989), Koch and Edstrom (2022) developed the Spiritual Harm & Abuse Scale (SHAS). These authors drew from the literature and interviews to find a wide range of experiences and emotional states that could be described as spiritual harm and abuse. From this work they developed a pool of 66 prompts for an online community survey which they fielded in 2021. Their study of the factor structure revealed a handful of factors which they used to distill items to a 27-item scale with six subscales.

In this paper, we build upon the work of Koch and Edstrom (2022) to further validate the SHAS by using item response theory. Item response theory can enhance clinical assessments in several ways. First, IRT provides interval-level measures which are more appropriate for statistical analysis than ordinal-level sum scores. Second, IRT improves upon sum scores by providing a scale along which both items and persons can be located. Third, measurements derived from IRT are sample-free, meaning it is possible to estimate a person's level of the latent construct free of the distribution of individual items and to estimate an item's difficulty level free from the distribution of people used in the sample. Finally, IRT improves upon single overall reliability coefficients by providing examination of precision across the score continuum.

The goal of this study was thus to use the Rasch Rating Scale Model (an IRT model) to enhance the SHAS in several ways. First, we aimed to express the SHAS as a single interval-level outcome measure of severity of spiritual harm and abuse. Second, we aimed to enrich the diagnostic value of the items and scale by examining severity. Third, we aimed to examine the distribution of item severities to see where the scale provides the most precise, reliable measurement of spiritual abuse. Such a scale also lays the foundation to further develop the SHAS by identifying anchor items that can calibrate with pilot items on future forms. Finally, we aimed to establish a SHAS score that is free of differential item functioning based on race, gender, and other demographic variables.

Methods

Item Pool Development

Koch and Edstrom (2022) developed an initial pool of items to measure spiritual harm and abuse. The authors validated the construct of spiritual harm and abuse through literature review (Kvarfordt, 2010; Nobakht & Dale, 2018; Oakley & Kinmond, 2014; Rodríguez-Carballeira, et al. 2015; Swindle, 2017; Ward, 2011, Winell, n.d.) and interviews with survivors of religious and spiritual abuse. These interviews revealed an array of specific examples of potential abuse not described in the published literature such as abuse inflicted by narcissistic persons, financial coercion, developmentally inappropriate children's teachings, pressure to stay in physically abusive marriages, neglect of needed medical care, sexual discrimination, and shunning/shaming, among others. The authors generated items by drawing from publicly available measures of similar constructs and by writing new items to maximize construct coverage. Each item they adapted to measure either a potentially spiritually abusive event (inflicting emotional abuse by a religious leader or group and/or with a religious or spiritual component, usually involving coercion or control) or a theorized effect of spiritual abuse, based on the reports of self-described survivors of spiritual abuse in the qualitative literature. In developing items, the authors sought to balance construct coverage with alternate wording and efficiency for participants. They wanted to cover as much of the domain as possible in a survey that could be completed in 15 minutes.

Participants and Procedures

Participants in the initial SHAS validation study were adults who responded to an online survey. To qualify for the survey, participants had to be at least 18 years of age and to have identified themselves as Christian in the past (Koch and Edstrom 2022).

The second author used snowball sampling to recruit participants to the study. He promoted the survey through various podcast feeds (in audio form during a podcast episode), e-mail lists, Facebook groups, and other online groups associated with progressive Christian podcasts based in the United States. Listeners/readers were encouraged to invite others who satisfied the criteria to take part in the survey. The second author also posted the link on his own social media accounts with a similar encouragement to invite others who satisfied the study criteria. The sampling priority was to reach the largest, most inclusive, community sample.

Participants were consented before beginning the survey. The study was approved by the IRB at Northwest University. Data collection began on January 28, 2021 and concluded on February 27, 2021.

Measures

Demographics

Demographic variables included gender, sexual orientation, race, raised in a Christian home, denomination, current religious identity, age, and theological self-description.

Spiritual Harm and Abuse Items

The final survey instrument included a total of 66 prompts separated into two distinct categories (Koch and Edstrom 2022). The first section, External Events, included 52 prompts about potentially abusive experiences. Prompts in this section measured prevalence by asking "How often have you experienced the above in a church or Christian group setting?" and providing participants a Liker scale of five response categories: (1) Never, (2) Once or twice, (3) Sometimes, (4) Often, and (5) All the time. The large number of prompts in this first section were presented in a randomized order so that all prompts would receive responses by approximately the same number of participants even if many participants did not complete all of the prompts in this section.

The second section, Internal States, included 14 prompts about personal feelings resulting from the abusive experiences described in the first section. Prompts in this section measured prevalence by asking "At any point, how often have you experienced the above as a result of negative religious experiences?" Participants were provided the same 5-point Likert scale of response categories as in the first section: (1) Never, (2) Once or twice, (3) Sometimes, (4) Often, and (5) All the time.

Data Analysis

In their initial validation of the SHAS, Koch and Edstrom (2022) examined the dimensionality of the 66 items. The authors used exploratory factor analysis (EFA) to determine the factor structure of the items. In these EFA analyses, the items loaded onto six inter-related factors of eigenvalues of at least 1. This enabled the authors to distill the item pool from 66 to 27 items and to identify several meaningful subscales for potential future use by clinicians and researchers.

For this study, we focused on the commonality of the items in order to use item response theory to establish a single interval scale of spiritual abuse. Item response theory models make two assumptions: (1) a single latent trait explains responses to items (unidimensionality) and (2) after controlling for this latent trait, items are weakly correlated (local independence) (Embretson and Reise 2000; Reise and Waller 2009; Reise, Bonifay, and Haviland 2013). The SHAS items were multidimensional by design. The construct of spiritual harm and abuse is not conceptually narrow, and to cover the breadth of this construct, Koch and Edstrom (2022) wrote two distinct item sets. However, a unidimensional IRT model can be appropriately fit to multidimensional data (Reise, Moore, and Haviland 2013). In this study we hypothesized a general factor of spiritual abuse as the primary cause of responses to the full set of 66 SHAS items. We therefore sought evidence of unidimensionality and local independence in order to calibrate the items and participants to a single IRT scale.

We began by examining classical item statistics to identify poorly functioning items based on extreme values of skew and kurtosis and low item-total correlations. Then we conducted several analyses of dimensionality in search of sufficient evidence of unidimensionality and local independence. This included principal components analysis to compare the eigenvalues of the largest and second largest components. Like factor analysis, component analysis attempts to summarize the common variance in items by identifying a smaller set of latent sources of covariation. Principal components analysis differs from factor analysis by treating all of the common variance among the items as the total variance to be explained [verify this]. We estimated a unidimensional Rasch model and calculated the proportion of variance in the data it explained. This we followed with a Principal Components Analysis of Residuals (PCAR) to gauge the importance of components in the item residuals, and with a similar statistic, Q3 ("Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model" 1984), to further inspect the correlations among item residuals.

We estimated a Rating Scale model [Andrich (1978)] to the SHAS items. We began by examining plots of the category characteristic curves of the items to visualize step parameters. Then we turned to item severity parameters and item fit statistics to inspect the technical quality of the items. We examined an item-person map to understand the fit of the items to the participants. We concluded by assessing the fit of the model to the data.

Finally, we conducted differential item functioning (DIF) analysis of the items to examine item bias by gender, and age, and race. [We applied logistic ordinal regression with IRT scoring. We used the Chi-squared likelihood-ratio statistic as the initial DIF detection criteria (alpha < 0.01), and a cut-off of McFadden pseudo R2Δ 0.02 in model comparisons to determine substantial DIF, a reasonable threshold used in the development of self-reported health outcomes.]

We conducted all analyses in **R** (R Core Team 2022). We used the **psych** (Revelle 2022) and **eRm** (Mair, Hatzinger, and Maier 2021) packages for analyses of dimensionality, the **mirt** (Chalmers 2022) and **TAM** (Robitzsch, Kiefer, and Wu 2022) packages for Rating Scale analyses, and the **lordif** (Choi, Laura E. Gibbons, and Crane 2016) package for DIF analyses.