



## Article

# Cross-Cultural Adaptation and Psychometric Properties of the SF-8 Questionnaire in Tanzanian Swahili for Injury Population

Leonardo Pestillo de Oliveira <sup>1,2,3</sup> , Nayara Malheiros Caruzzo <sup>2,3</sup> , Francis Sakita <sup>4</sup>, Blandina T. Mmbaga <sup>3,4,5,6</sup>, Judith Boshe <sup>4</sup>, Anna Tupetz <sup>7</sup>, Catherine A. Staton <sup>3,7</sup> and Joao Ricardo Nickenig Vissoci <sup>2,3,7,\*</sup>

<sup>1</sup> Graduate Program of Health Promotion, Cesumar Institute of Science, Technology and Innovation, Cesumar University, Maringá 87050-900, Brazil; leopestillo@gmail.com

<sup>2</sup> Associate Postgraduate Program in Physical Education UEM/UEL, State University of Maringá, Maringá 87020-900, Brazil; nayaramalheiros@gmail.com

<sup>3</sup> Duke Global Health Institute, Duke University, Durham, NC 27710, USA; b.mmbaga@kcri.ac.tz (B.T.M.); catherine.lynch@duke.edu (C.A.S.)

<sup>4</sup> Kilimanjaro Christian Medical Centre, Moshi, Tanzania; francis.sakita@gmail.com (F.S.); jboshe@live.com (J.B.)

<sup>5</sup> Clinical Trial Department, Kilimanjaro Clinical Research Institute, Moshi, Tanzania

<sup>6</sup> Pediatric and Child Health Department, Kilimanjaro Christian Medical University College, Moshi, Tanzania

<sup>7</sup> Duke Emergency Medicine, Duke University Medical Center, Durham, NC 27710, USA; anna.tupetz@duke.edu

\* Correspondence: jnv4@duke.edu



**Citation:** Pestillo de Oliveira, L.; Malheiros Caruzzo, N.; Sakita, F.; Mmbaga, B.T.; Boshe, J.; Tupetz, A.; Staton, C.A.; Nickenig Vissoci, J.R. Cross-Cultural Adaptation and Psychometric Properties of the SF-8 Questionnaire in Tanzanian Swahili for Injury Population. *Disabilities* **2022**, *2*, 428–438. <https://doi.org/10.3390/disabilities2030030>

Academic Editor: Benjamin J. Lovett

Received: 29 April 2022

Accepted: 7 July 2022

Published: 23 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** *Background:* There is a lack of tools to screen for health-related quality of life (HRQoL) in acute injury patients, despite the critical need for having a good understanding of the characteristics of mental health during the rehabilitation process. The SF-8 instrument, a shorter version of the SF-36, is the most widely used patient-based assessment of HRQoL. The aim of this research is to adapt the psychometric properties of the SF-8 to Swahili. *Methods:* This study is a secondary data analysis of previously collected and psychometric evaluation of the culturally adapted and translated SF-8. A cross-cultural adaptation committee carried out the process of translation to provide validity evidence based on test content. Confirmatory factor analysis was used to test the internal structure-based evidence. The validity based on relation to other variables (discriminant evidence) was tested using polychoric correlation with PHQ-2 (Patient Health Questionnaire-2). The reliability was tested using Cronbach's alpha, Omega McDonald, and Composite Reliability. *Results:* 1434 adults who suffered an acute injury and presented to the emergency department between April 2018 and August 2020 were included in the study. The instrument demonstrated language clarity and domain coherence, showing validity evidence based on test content. The CFA (Confirmatory Factor Analysis) analysis showed good fit indices for both models (one- and two-factor models) of the SF-8. The discriminant evidence showed that SF-8 scores correlate strongly with the PHQ-2 instrument. These results supported the validity evidence in relation to other variables. All analyses of reliability were considered adequate with values above 0.90 for both models of the SF-8. *Conclusions:* The results show that the SF-8 instrument can provide relevant information about the health-related quality of life of acute injury patients, and allow practitioners to gain a better understanding of mental health, improving the treatment and follow-up of injury patients within Tanzanian culture.

**Keywords:** validation; SF-8; quality of life; Tanzania; Swahili; psychometric properties; injury; rehabilitation; disability

## 1. Introduction

Health-related quality of life (HRQoL) is a multidimensional construct encompassing patients' perspectives on the effect a disease or medical treatment has on their general well-being [1]. As such, HRQoL includes social, physical, and emotional functions [1,2]. Guidelines for clinical trials, pharmaceutical development, and patient advocacy have

stressed the value of assessing HRQoL to guide clinical practice and medical research [3,4]. The ability to adequately evaluate health-related quality of life is a critical need in the development of acute injury research. In fact, to quantify the subjective burden of a person's experienced trauma, it is essential to understand the injury survivors' health-related quality of life. Short and long-term disabilities can result in a variety of functional, physical, emotional, cognitive, and social challenges and thereby directly impact a person's quality of life (QoL). To appropriately care for these patients, having an instrument that allows a multifaceted assessment of these areas is critical [5]. Patient-reported outcomes assessing HRQoL allow us to understand the patient's perceived level of dysfunction specific to their individual social context and inform a patient-centered individualized assessment [6].

Especially in Tanzania, where no such instrument has ever been validated, there is a pressing need for reliable objective measures to inform practice and policy. Low- and middle-income countries account for 94% of all injury-related disability-adjusted life years (DALYs), a composite score of years lived with a disability (YLD) and years of life lost (YLL) [7]. In Tanzania, specifically, it is estimated that around 13% of all years lived with a disability (YLD) are attributed to injuries [8].

Polinder et al. found that the SF-36 was most frequently used to assess the quality of life in patients after traumatic brain injuries (TBI) [5]. Among the studies analyzed, those who performed the validation of the instrument presented positive results about the internal consistency and interpretability of the instrument. The meta-analysis showed that TBI can be considered a heterogeneous condition that covers a broad spectrum of HRQoL. Yet, there is a lack of validated QoL assessment tools in the general injury population, especially in low- and middle-income country (LMIC) settings.

The SF-8 is a shorter version of the SF-36, the most widely used patient-based assessment of the health-related quality of life [9]. The SF-36 instrument can be used to estimate a single-index measure for health based on the general population, and this is an advantage of its use [5]. In the same way, the SF-8 instrument has an advantage mainly related to its brevity: its score can be compared with that of the longer versions, and administration takes one to two minutes [10]. Further research is needed to provide more evidence of internal structure validity for the SF-8 in an injury sample.

Intended for use as a screening or monitoring instrument in large population surveys, the SF-8 was originally developed in the U.S. and has been translated to over 40 languages [9,10]. Following its initial validation, a few studies have assessed the validity of the SF-8 in American migraineurs, Spanish surgical patients, the Japanese general population, and conflict-affected populations in northern Uganda [2,9,11,12]. However, cross-cultural studies have not yet been conducted an assessment of the validity of a Swahili version of the SF-8 in a Tanzanian setting. Additionally, only one study has demonstrated the use of the SF-8 as a measure of patient-reported outcome specifically in patients with TBI [13]. A reliable and validated version of the SF-8 in Swahili will improve the ability to conduct quality of life studies in Tanzania with injury patients. Thus, the aim of the present study is to report on the psychometric properties included in the first translation and the adaptation of the SF-8 to Swahili, specifically looking at (a) translating and adaptation, (b) verifying internal structure and consistency, and (c) providing evidence of validity.

## 2. Materials and Methods

This study is a secondary data analysis of a previously conducted psychometric evaluation of the culturally adapted and translated SF-8. The underlying data source is a trauma registry in northern Tanzania, which is part of a study protocol that was approved by the Duke University Medical Center Institutional Review Board (Pro00086496), the Kilimanjaro Christian Medical Center (KCMC) Ethics Committee, as well as the Tanzanian National Institute of Medical Research (NIMR). The collected registry data are part of an ongoing quality improvement process and do not require informed consent by the included injury patients, as approved by KCMC and NIMR.

### 2.1. Participants

The study sample is composed of 1434 adults who were injury patients from a trauma registry in northern Tanzania between April 2018 and August 2020 (Table 1). The inclusion criteria were: being at least 18 years of age, being able to speak Swahili, being able to understand and respond appropriately to the questionnaires, and having given consent to participate prior to hospital discharge. The attending physician evaluated the ability to understand and respond appropriately to self-reported scales, such as the SF-8. Only those patients who were deemed cognitively able by the physician were approached. Participants were patients admitted to KCMC Emergency Department (ED) for management of acute injury (<24 h). Patients referred from another hospital were also included in the registry if the time from injury was below 24 h.

**Table 1.** Sociodemographic profile of the validation sample.

Characteristic	N	N = 1434
<b>Age</b> <sup>1</sup>	1434	38.44 (±16.37)
<b>Sex</b>	1427	
Male		1162 (81%)
Female		265 (19%)
<b>Marital status</b>	1422	
Married		763 (54%)
Not married		659 (46%)
<b>Occupation</b>	1421	
Self-employed		926 (65%)
Employed		339 (24%)
Other		102 (7%)
Unemployed		54 (4%)
<b>Education</b>	1401	
Primary		891 (64%)
Secondary		407 (29%)
University		103 (7%)
<b>Mechanism of injury</b>	1433	
Road traffic injury		909 (63%)
Fall		237 (17%)
Assault		198 (14%)
Other		89 (6%)
<b>Injury severity</b>	1252	
Mild		986 (79%)
Moderate		219 (17%)
Severe		47 (4%)
<b>Traumatic Brain Injury</b>	825	260 (32%)

<sup>1</sup> Average (SD).

### 2.2. Instrument

The original scale used was SF-8, developed by QualityMetric [10]. The SF-8 is a quickly administered patient-reported outcome measure that can be used to assess health-related quality of life. The instrument consists of eight single-item scales addressing eight domains of (1) general health, (2) physical functioning, (3) role limitations due to physical health problems, (4) bodily pain, (5) vitality (energy / fatigue), (6) social functioning, (7) mental health, and (8) role limitations due to emotional problems. Each answer is scored on a six (items 1 and 4) or five-point (items 2, 3, 5, 6, 7, and 8) Likert type scale [10].

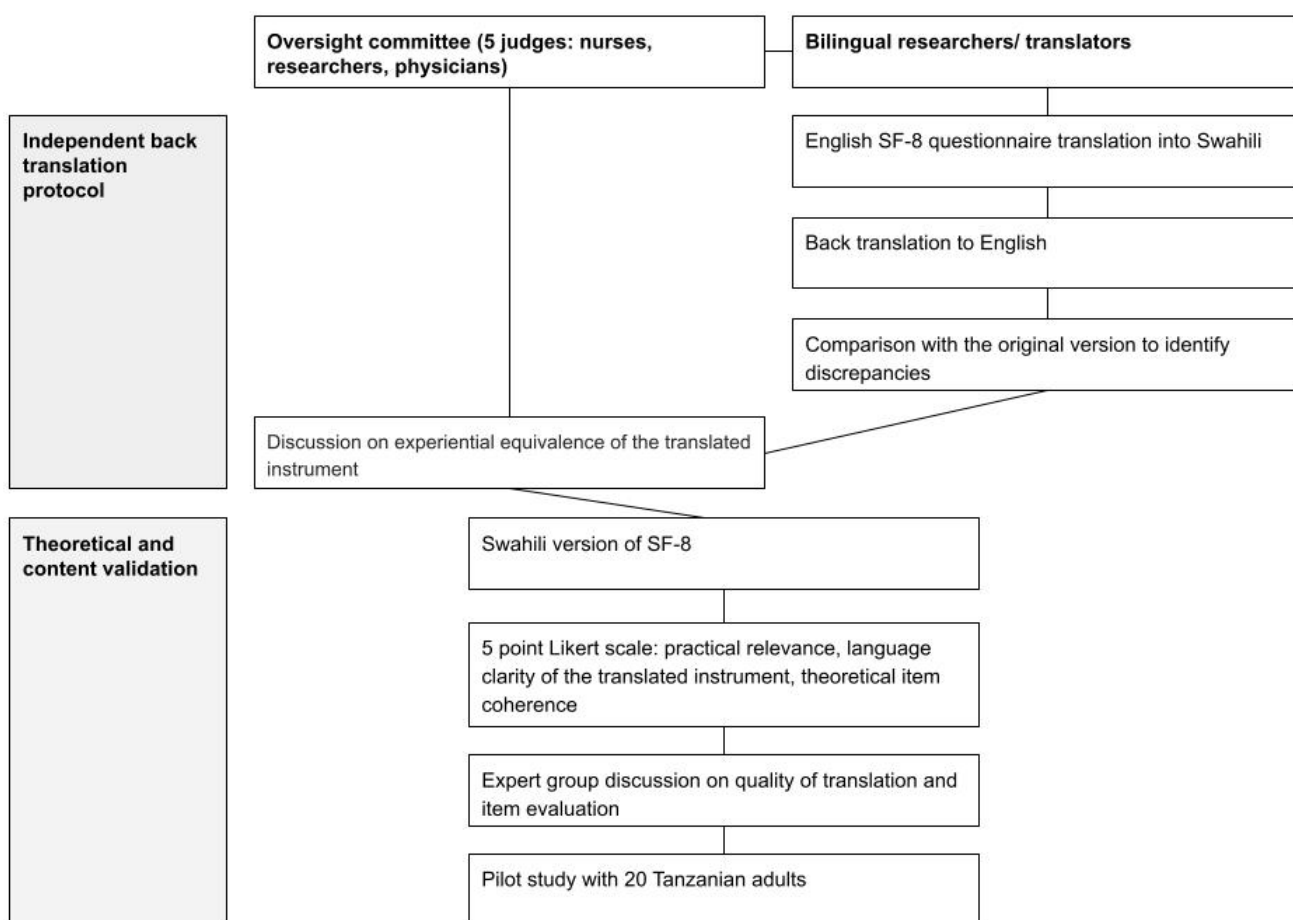
### 2.3. Translation and Cross-Cultural Adaptation

We formed a translation and cross-cultural adaptation committee of five judges (physicians, nurses, and researchers) who oversaw the translation, adaptation, and the process of

validity based on the test content. The SF-8 instrument translation was performed following our independent back translation protocol. This protocol presents the following steps: (a) a Swahili translator translated the SF-8 questionnaire into Swahili; (b) a bilingual translator translated the Swahili version back into English; (c) a comparison of the English-translated version with the original version of the instrument and screening for inconsistencies by four independent bilingual researchers, and, lastly, (d) issues with experiential equivalence were detected and discussed by the researchers and the juror’s committee during the translation meetings [14]. These data were not formally reported, but a consensus was reached and adjustments were made to find appropriate expressions.

To perform a theoretical and content evaluation of the translated instrument, we used a five-point Likert scale created to verify the practical relevance, the language clarity of the translated instrument, and the theoretical coherence of the item (Figure 1).

**Translation and cross-cultural adaptation SF-8**



**Figure 1.** Processes of translation and cross-cultural adaptation.

After finalizing content validation, we conducted a small pilot study with a group of 20 Tanzanian adults at KCMC, selected by convenience, to confirm the quality of instrument questions and coherence of language and content.

**2.4. Data Collection**

Patients in the Kilimanjaro Christian Medical Center injury registry were screened for inclusion in this project. They were offered enrollment prior to discharge from the hospital. The patient injury severity was assessed by a modified Kampala Trauma Score II, in which the “number of serious injuries” component was dichotomized into either severely injured or not. The presence of TBI, need for surgery, polytrauma, spine or long

bone fractures, and Operation Room (OR) or Intensive Care Unit (ICU) placement was used for the classification. The patient had the SF-8 questions administered at the bedside as a part of a 45-min interview. All responses were collected by hand and were entered into an Internet-based dataset (REDCap); all data were assessed for quality at collection, and data entry. Additionally, the data underwent a final quality check by the principal investigator (PI).

### 2.5. Data Analysis

This study provided validity evidence of the SF-8 in Tanzania based on the sources of validity evidence described in the Standards for Educational and Psychological Testing [15]. The validity evidence examined included test content, internal structure, and relation to the other variables. We did not provide validity evidence based on response processes or with respect to the consequences of testing. The SF-8 questionnaire was collected by the nurse, as part of an interview, and these data available were not available.

Sociodemographic data were presented as means and standard deviations or absolute and relative frequencies. All analyses were conducted with R Language for Statistical Computing (version 4.0.2 for Mac, R Core Team, Vienna, Austria) [16]. Validity and reliability analyses were performed following Anthoine et al.'s review of patient-reported outcome scale validations [17]. Descriptive statistics were used to present the participant demographics and injury characteristics of our sample.

### 2.6. Evidence of Validity

#### 2.6.1. Evidence of Validity Based on Content

Haynes et al. define content validity as the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose [18]. Based on "Standards for Educational and Psychological Testing", it is "an analysis of the relationship between the content of a test and the construct it is intended to measure" [15]. In this study, validity based on content was evaluated by group discussions among multiple participants about a specified topic of interest. In this case, experts in studies about functional performance of brain-injured patients and global health participated. They responded to a Likert-type scale about the clarity of language and theoretical and practical pertinence of each item of the scale. To analyze the concordance index between judges for the theoretical dimensions of the items, the content validity coefficient (CVC) was used.

#### 2.6.2. Evidence of Validity Based on Internal Structure

The internal structure-based validity was examined with Confirmatory Factor Analysis (CFA) using R packages "lavaan" (version 0.6.12) and "semPlot" (version 1.1.5) to evaluate the goodness-of-fit of the unidimensional and two-factor model to the data (Table 2). CFA was used to verify the construct validity of the instrument through (a) item-factor parameter and items' individual reliability; (b) absolute, incremental, and parsimonious fit indexes; and (c) average variance extracted to examine the validity [19,20].

CFA model adequacy was tested using Weighted Least Square Means and Variance Adjusted (WLSMV). Model adjustment was tested through the fit indices (reference of expected values for each index): Chi-square ( $\chi^2$  and  $p$ -value), Root Mean Square Error of Approximation (RMSEA < 0.08, CI 90%); Normalized Fit Index (NFI > 0.90), Tucker–Lewis index (TLI > 0.90) and Comparative Fit Index (CFI > 0.90). These indices aim to assess whether the model is a good fit for the data, as proposed in the literature [20,21]. Average Variance Extracted (AVE) was evaluated with values higher than 0.50 considered acceptable indicators [22]. The Composite Reliability (CR) was calculated using CFA results, as the values that this measure provides are important to know the index of internal consistency of the dimensions of the instrument through the factorial loads of the respective items. If the values are greater than 0.70 they are considered indicators of suitable composite reliability [23]. Factor loading and thresholds are reported.

**Table 2.** Reliability and confirmatory factor analysis model fit indicators.

	SF-8 1 Factor Model	SF-8 2 Factor Model
<b>Reliability</b>		
Cronbach's Alpha (CI 95%)	0.971 (0.968, 0.973)	Physical = 0.964 (0.960, 0.966) Mental = 0.933 (0.926, 0.938)
Omega McDonald	0.958	Physical = 0.940 Mental = 0.839
Composite Reliability	0.968	Physical = 0.979 Mental = 0.941
<b>CFA</b>		
$\chi^2$ (df)/p-value	261.070 (20)/0.000	226.599 (19)/0.000
RMSEA (CI 95%)	0.092 (0.082, 0.100)	0.087 (0.077, 0.098)
TLI	0.989	0.990
CFI	0.992	0.993
NFI	0.989	0.990
Average extracted variance	0.841	Physical = 0.888 Mental = 0.833

CFA = Confirmatory Factor Analysis;  $\chi^2$  = Chi-Square; df = Degree of Freedom; RMSEA = Root Mean Square Error of Approximation; TLI = Tucker-Lewis Index; CFI = Comparative Fit Index.

### 2.6.3. Evidence of Validity Based on Relation to Other Variables

Evidence of validity based on relation to other variables is usually established by a high correlation between the scale and different measures (discriminant evidence) or similar constructs (convergent evidence). In this case, we assessed the discriminant evidence by the polychoric correlation between PHQ-9 (Patient Health Questionnaire-9) and SF-8 scale. The PHQ-9 is a self-administered questionnaire, previously validated in the Tanzanian context, based on the clinician-administered Primary Care Evaluation of Mental Disorders (PRIME-MD). It is built on the nine criteria used to diagnose depression according to the Diagnostic and Statistical Manual of Mental Disorders 4th ed. (DSM-IV) [24,25]. We used the items 1 ("In the past two weeks how often have you been bothered by little interest or pleasure in doing things") and 2 ("In the past two weeks how often have you been bothered by feeling down, depressed, or hopeless") in our analysis. The abbreviated PHQ-2 has previously been validated as a sensitive tool for depression screening [26].

### 2.7. Multigroup Invariance Analysis

We used multigroup Confirmatory Factor Analysis to measure the invariance for the groups. This analysis allows us to evaluate whether the configuration and parameters of the instrument are invariant (equivalent) for different groups of people. In the present study, we carried out the multigroup analysis to investigate the use of the scale in different sexes (male and female) and ages (<35 and >35 years old). The invariance is observed through  $\Delta$ CFI, where a variation <0.01 is expected in groups >300 subjects (>35 = 774; <35 = 659) or <0.03 when groups are <300 (male = 1161, female = 265) [27].

### 2.8. Evidence of Reliability

Reliability refers to the capacity of an instrument to produce reliable results in different situations. Reliability with internal consistency was measured to ensure all of the items in the instrument refer to the same subject [28]. In order to measure reliability, we used a set of scores (Cronbach's alpha, Omega McDonald, and Composite Reliability), in which coefficients above 0.7 were considered acceptable [20,29].

### 3. Results

#### 3.1. Sample Characteristics

Most of the population was male (81%), married (54%), and had primary education (64%). Furthermore, 29% of the study sample had secondary education and 7.4% had university education. The average age (SD) was 38.44 ( $\pm 16.37$ ) (Table 1). The majority of participants were self-employed (36%) and suffered a road traffic injury (64%). Mild/moderate injuries represented roughly 96% of cases as classified by the modified KTS-II scale applied.

#### 3.2. Evidence of Validity Based on Content

All coefficients obtained by CVC analysis related to language clarity and domain coherence were above 0.80. These findings indicate that the translated and adapted version of the SF-8 questionnaire are clearly understandable within Tanzanian culture, in addition to being relevant and pertinent. SF-8 item classification agreement among judges was also above 0.80, indicating that the evaluators found the items to be consistent with the underlying theoretical conceptualization (Appendix A).

#### 3.3. Evidence of Validity Based on Internal Structure

As reported in previous studies [2,11,12], we tested two versions of SF-8 (unidimensional and two-factors models). CFA results indicated that the uni-dimensional ( $\chi^2 = 261.070$ ,  $df = 20$ ,  $p < 0.001$ ; CFI = 0.992; RMSEA = 0.092 (0.082–0.100) and TLI = 0.989) and two-factor model of SF-8 ( $\chi^2 = 226.599$ ,  $df = 19$ ,  $p < 0.001$ ; CFI = 0.993; RMSEA = 0.087 (0.077–0.098) and TLI = 0.990), presented an acceptable fit for the data. Both models performed well and the fit indices suggested that a unidimensional model and a two-factor model were adequate for SF-8 (Table 2). Factor loadings were between 0.86 and 0.99 for both models (Figure 2); AVE for the unidimensional model was 0.84 and for the two-factors model the AVE was 0.89 and 0.83 (Physical and Mental dimension, respectively) (Table 2). Both models showed a significant chi-square  $p$ -value, which does not match the recommendation values for the fit indices. It is reported in the literature that  $\chi^2$  is very sensitive to the sample size and there is a strong possibility that  $\chi^2$  is significant even when the model presents a good fit. Because voluminous samples are required to perform CFA, other indices are commonly used to reach more confidence [22]. The fit indices provide validity evidence based on the internal structure of this instrument.

#### 3.4. Reliability Evidence

Reliability was considered adequate with values above 0.90 for SF-8 in all reliability measurements, indicating strong internal consistency for both models of the SF-8 scale (Table 2).

#### 3.5. Evidence of Validity Based on Relation to Other Variables

Polychoric correlation coefficients revealed that the SF-8 scores were statistically ( $p < 0.001$ ) and positively correlated with items 1 and 2 of PHQ-9 scale ( $r = 0.490$  and  $r = 0.450$ , respectively for one-factor model and  $r = 0.509$ ;  $r = 0.403$  for Physical and  $r = 0.472$ ;  $r = 0.492$  for the Mental dimension of the two-factor model of SF-8 (Table 3). These statistically significant and substantial relationships provided validity evidence based on relations to other variables.

#### 3.6. Multi Groups Invariance Evidence

Tables 4 and 5 show that there was invariance for sex and age groups indicating that both models of the SF-8 are equivalent among those groups.

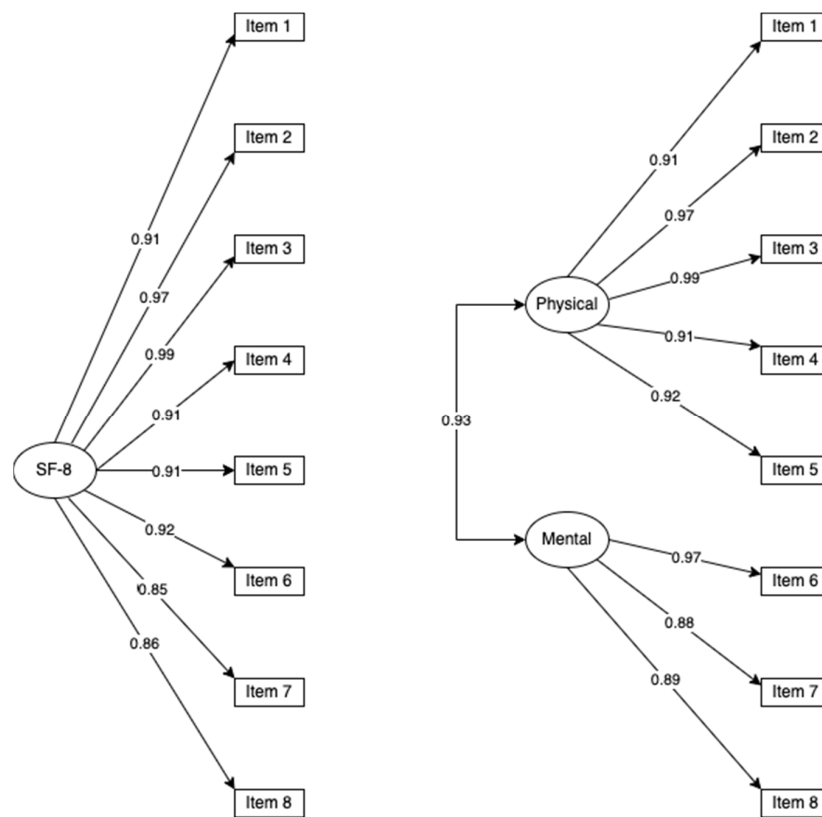


Figure 2. Confirmatory factor analysis diagram, factor loadings for the SF-8.

Table 3. Polychoric correlation coefficients of SF-8 items and PHQ9 items.

SF-8 Items	PHQ Item 1	PHQ Item 2
SF-1	0.598	0.532
SF-2	0.615	0.638
SF-3	0.697	0.677
SF-4	0.652	0.631
SF-5	0.573	0.530
SF-6	0.681	0.716
SF-7	0.685	0.718
SF-8	0.727	0.750
1-factor model	0.490	0.450
2-factors Physical	0.509	0.403
2-factors Mental	0.472	0.492

Table 4. Invariance of the confirmatory model by sex and age group for the unidimensional model of SF-8.

	Invariance	$\chi^2$	df	RMSEA (90% CI)	CFI	$\Delta$ CFI	TLI
By sex	Form	293.225	40	0.094 (0.084–0.100)	0.991		0.988
	Loadings	214.060	47	0.071 (0.061–0.080)	0.994	0.001	0.993
	Variance	211.273	46	0.071 (0.061–0.081)	0.994	>0.001	0.993
By age	Form	261.224	40	0.088 (0.078–0.098)	0.992		0.989
	Loadings	200.747	47	0.068 (0.058–0.077)	0.995	0.003	0.994
	Variance	198.102	46	0.068 (0.058–0.078)	0.995	>0.001	0.994



**Table 5.** Invariance of the confirmatory model by sex and age group for a two-factors model of SF-8.

	Invariance	$\chi^2$	df	RMSEA (90% CI)	CFI	$\Delta$ CFI	TLI
By sex	Form	254.507	38	0.089 (0.079–0.100)	0.993		0.989
	Loadings	203.989	44	0.071 (0.062–0.082)	0.995	0.002	0.993
	Variance	198.311	42	0.072 (0.062–0.083)	0.995	>0.001	0.993
By age	Form	223.189	38	0.083 (0.072–0.093)	0.994		0.991
	Loadings	195.686	44	0.069 (0.060–0.079)	0.995	0.001	0.993
	Variance	190.199	42	0.070 (0.060–0.081)	0.995	>0.001	0.993

#### 4. Discussion

This is the first study to conduct a cross-cultural validation of the SF-8 scale in Swahili in Tanzania. Moreover, this is also the first study to show evidence of the psychometric properties of this scale in injury patients from Tanzania. Most psychometric properties reports have been conducted with general populations, migraineurs, and surgical patients [2,9,11]. The Tanzanian version of the SF-8 showed adequate evidence based on test content, and in relation to other variables. Internal structure showed adequate fit in some indicators, but limitations in model fit were also identified. Our results present a preliminarily cross-cultural adaptation and validation of the SF-8.

The translated and adapted SF-8 version showed similar evidence based on test content values to the ones found in the literature [2,9,11,12]. Previous psychometric analyses have shown SF-8 internal structure as a unidimensional scale [11].

However, some other studies have only shown the two-factor models [2,12]. In our analysis of the internal structure of the Swahili version of SF-8, the analysis of one- and two-dimension models demonstrated a good fit on CFI and TLI, but had significant chi-square tests and high intervals in RMSEA, suggesting that any of these two models are acceptable. Reliability scores of both versions (one- and two-dimension models) met internal consistency criteria as suggested in the literature, being equal or higher than 0.70 [21], and were also consistent with previous studies [2].

Discriminant evidence has been demonstrated between the SF-8 and PHQ-2. Our results confirm the instrument's ability to perform as a theoretical concept of HRQoL including social, physical, and emotional function constructs.

One limitation with this study is related to its sample, which consisted of injury patients enrolled at a single center in Tanzania. While this sample provides an initial assessment of the psychometric properties of the SF-8 Swahili version, it offers a limited perspective on the general Tanzanian population. Furthermore, traumatic injury disproportionately affects males, in Tanzania and globally, which was reflected in our imbalanced sample. Additionally, our multigroup analysis showed invariance for gender, as well as for the age groups. Future studies should approach these issues concerning cross-validation such a way that the results may be generalized to other independent samples in Tanzania, patient populations, and cultures to confirm the factor solution stability found in our results, specifically with confirmatory approaches.

However, our study findings from a Tanzanian injury sample can now be reproduced in a larger population. Our results demonstrate that the SF-8 can provide relevant information to elucidate health-related quality of life amongst injury patients, allowing practitioners to understand different dimensions and potentially giving parameters to the assessment and follow-up of this population group.

Another limitation was found at the CFA analysis, where the chi-square/df values are higher than the recommended for both models (1 and 2 factors). Kline (2010) [20] highlights that, commonly, values of statistical tests in SEM (Structural Equation Modeling) analysis tend to be too high. One of the reasons is because of the distributional characteristics of the data. Additionally, the p values for test statistics are also sensitive to sample size and on sampling distributions. Since our samples were collected at a hospital where cases were not selected using a chance-based method, additional studies with other samples should be

conducted, to confirm if the structure of the SF-8 has limitations due to the hospital based population or another structure of the instrument. Exploratory analysis could also be used. Moving forward, we recommend that other psychometric properties should be addressed, such as responsiveness or even individual item parameters.

Since this is the first instrument to be validated in Swahili, more research is needed to establish strong evidence of SF-8 behavior in relation to how the studied HRQoL components correlate with other HRQoL measures.

## 5. Conclusions

In summary, the scale has shown good psychometric properties and could be used as a unidimensional or as a two-factor scale to assess HRQoL in Tanzanian culture. This is an initial report on the psychometric properties of the SF-8 to lead to further capacity development to measure and improve HRQoL in Tanzania, allowing for more evidence-based practices and advancements in research and policy development.

**Author Contributions:** Data analysis, L.P.d.O., N.M.C. and J.R.N.V.; writing, L.P.d.O., N.M.C. and A.T.; visualization, N.M.C. and A.T.; supervision, F.S., B.T.M., J.B., C.A.S. and J.R.N.V.; editing, F.S., B.T.M., J.B., A.T. and J.R.N.V.; conceptualization, F.S., B.T.M., J.B., C.A.S. and J.R.N.V.; funding acquisition, B.T.M. and C.A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** Funding for this project was received from the Fogarty International Center of the National Institutes of Health under Award Number 1R21TW010456-01A1 (PI, Staton).

**Institutional Review Board Statement:** This study was approved by the Duke University Health System Institutional Review Board (Pro00061652 and Pro00086496), the Kilimanjaro Christian Medical University College Ethics Committee (Certificate #821 and #1087), and the Tanzania National Institute for Medical Research (NIMR), Tanzania.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. This manuscript is a secondary data analysis of previously collected data.

**Data Availability Statement:** The data presented in this study are available on reasonable request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Content Validity Coefficient (CVC) Scores per SF-8 Item.

SF-8 Items	CVC Score
SF-1	0.87
SF-2	0.87
SF-3	0.93
SF-4	0.80
SF-5	0.93
SF-6	0.80
SF-7	0.80
SF-8	0.80

## References

1. Revicki, D.A.; Kleinman, L.; Cella, D. A history of health-related quality of life outcomes in psychiatry. *Dialogues Clin. Neurosci.* **2014**, *16*, 127–135. [[CrossRef](#)] [[PubMed](#)]
2. Vallès, J.; Guilera, M.; Briones, Z.; Gomar, C.; Canet, J.; Alonso, J.; ARISCAT Group. Validity of the Spanish 8-item short-form generic health-related quality-of-life questionnaire in surgical patients: A population-based study. *Anesthesiology* **2010**, *112*, 1164–1174. [[CrossRef](#)] [[PubMed](#)]
3. Beitz, J. Quality-of-life end points in oncology drug trials. *Oncology* **1999**, *13*, 1439–1442. [[PubMed](#)]

4. Badia, X.; Herdman, M. The importance of health-related quality-of-life data in determining the value of drug therapy. *Clin. Ther.* **2001**, *23*, 168–175. [[CrossRef](#)]
5. Polinder, S.; Haagsma, J.A.; van Klaveren, D.; Steyerberg, E.W.; van Beeck, E.F. Health-related quality of life after TBI: A systematic review of study design, instruments, measurement properties, and outcome. *Popul. Health Metr.* **2015**, *13*, 4. [[CrossRef](#)] [[PubMed](#)]
6. Fairbairn, K.; May, K.; Yang, Y.; Balasundar, S.; Hefford, C.; Abbott, J.H. Mapping Patient-Specific Functional Scale (PSFS) Items to the International Classification of Functioning, Disability and Health (ICF). *Phys. Ther.* **2012**, *92*, 310–317. [[CrossRef](#)] [[PubMed](#)]
7. Chandran, A.; Hyder, A.A.; Peek-Asa, C. The Global Burden of Unintentional Injuries and an Agenda for Progress. *Epidemiol. Rev.* **2010**, *32*, 110–120. [[CrossRef](#)] [[PubMed](#)]
8. Institute for Health Metrics and Evaluation. GBD Results Tool | GHDx [Internet]. *GBD Results Tool*. 2021. Available online: <http://ghdx.healthdata.org/gbd-results-tool> (accessed on 23 March 2021).
9. Turner-Bowker, D.M.; Bayliss, M.S.; Ware, J.E., Jr.; Kosinski, M. Usefulness of the SF-8 Health Survey for comparing the impact of migraine and other conditions. *Qual. Life Res.* **2003**, *12*, 1003–1012. [[CrossRef](#)] [[PubMed](#)]
10. Ware, J.E.; Kosinski, M.; Dewey, J.; Gandek, B. *How to Score and Interpret Single-Item Health Status Measures: A Manual for Users of the SF-8 Health Survey*; QualityMetric Inc.: Boston, MA, USA, 2001.
11. Tokuda, Y.; Okubo, T.; Ohde, S.; Jacobs, J.; Takahashi, O.; Omata, F.; Yanai, H.; Hinohara, S.; Fukui, T. Assessing Items on the SF-8 Japanese Version for Health-Related Quality of Life: A Psychometric Analysis Based on the Nominal Categories Model of Item Response Theory. *Value Health* **2009**, *12*, 568–573. [[CrossRef](#)] [[PubMed](#)]
12. Roberts, B.; Browne, J.; Ocaka, K.F.; Oyok, T.; Sondorp, E. The reliability and validity of the SF-8 with a conflict-affected population in northern Uganda. *Health Qual. Life Outcomes* **2008**, *6*, 108. [[CrossRef](#)] [[PubMed](#)]
13. Sharma, A.; Sane, H.; Kulkarni, P.; Yadav, J.; Gokulchandran, N.; Biju, H.; Badhe, P. Cell therapy attempted as a novel approach for chronic traumatic brain injury—A pilot study. *SpringerPlus* **2015**, *4*, 26. [[CrossRef](#)] [[PubMed](#)]
14. Cha, E.-S.; Kim, K.H.; Erlen, J.A. Translation of scales in cross-cultural research: Issues and techniques. *J. Adv. Nurs.* **2007**, *58*, 386–395. [[CrossRef](#)] [[PubMed](#)]
15. American Educational Research Association, Association AP, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*; Amer Educational Research Assn: Washington, DC, USA, 2014.
16. R: The R Project for Statistical Computing. 2021. Available online: <https://www.r-project.org/> (accessed on 15 December 2021).
17. Anthoine, E.; Moret, L.; Regnault, A.; Sébille, V.; Hardouin, J.-B. Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. *Health Qual. Life Outcomes* **2014**, *12*, 176. [[CrossRef](#)] [[PubMed](#)]
18. Hernández-Nieto, R. *Contributions to Statistical Analysis: The Coefficients of Proportional Variance, Content Validity and Kappa*; BookSurge Publishing: Charleston, SC, USA, 2002.
19. DeVellis, R.F. *Scale Development: Theory and Applications*, 2nd ed.; Sage Publications: Thousand Oaks, CA, USA, 2003.
20. Kline, R.B. *Principles and Practice of Structural Equation Modeling*, 3rd ed.; Methodology in the Social Sciences; The Guilford Press: New York, NY, USA, 2010.
21. Byrne, B.M. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*; Routledge: London, UK, 2013.
22. Hair, J.F.; Babin, B.; Anderson, R.E.; Tatham, R.L. SEM: Confirmatory Factor Analysis. In *Multivariate Data Analysis*, 6th ed.; Pearson Prentice Hall: Hoboken, NJ, USA, 2006; pp. 770–836.
23. Nunnally, J.; Bernstein, I. *Psychometric Theory*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1994.
24. Kroenke, K.; Spitzer, R.L.; Williams, J.B. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* **2001**, *16*, 606–613. [[CrossRef](#)] [[PubMed](#)]
25. Fawzi, M.C.S.; Ngakongwa, F.; Liu, Y.; Rutayuga, T.; Siril, H.; Somba, M.; Kaaya, S.F. Validating the Patient Health Questionnaire-9 (PHQ-9) for screening of depression in Tanzania. *Neurol. Psychiatry Brain Res.* **2018**, *31*, 9–14. [[CrossRef](#)] [[PubMed](#)]
26. Kroenke, K.; Spitzer, R.L.; Williams, J.B. The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Med. Care* **2001**, *41*, 1284–1292. [[CrossRef](#)] [[PubMed](#)]
27. Svetina, D.; Rutkowski, L.; Rutkowski, D. Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using M plus and the lavaan/semTools Packages. *Struct. Equ. Model. A Multidiscip. J.* **2019**, *27*, 1–20. [[CrossRef](#)]
28. Roach, K.E. Measurement of Health Outcomes: Reliability, Validity and Responsiveness. *JPO J. Prosthet. Orthot.* **2006**, *18*, 8–12. [[CrossRef](#)]
29. Lohr, K.N.; Aaronson, N.K.; Alonso, J.; Burnam, M.A.; Patrick, D.L.; Perrin, E.B.; Roberts, J.S. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clin. Ther.* **1996**, *18*, 979–992. [[CrossRef](#)]