

## Research Article

**Cite this article:** Keyan D, Hadzi-Pavlovic D, Akhtar A, Dawson K, Koyiet PN and Bryant R (2024) Dimensionality of the Swahili version of the General Health Questionnaire (GHQ-12) in a Kenyan population: A confirmatory factor analysis. *Cambridge Prisms: Global Mental Health*, **11**, e48, 1–9  
<https://doi.org/10.1017/gmh.2024.46>.

Received: 06 February 2023

Revised: 05 December 2023

Accepted: 03 April 2024

### Keywords:




confirmatory factor analysis; factor structure; GHQ-12; Kenya; mental health disorder

### Corresponding author:

Dharani Keyan;

Email: [d.keyan@unsw.edu.au](mailto:d.keyan@unsw.edu.au)

# Dimensionality of the Swahili version of the General Health Questionnaire (GHQ-12) in a Kenyan population: A confirmatory factor analysis

Dharani Keyan<sup>1</sup> , Dusan Hadzi-Pavlovic<sup>2</sup>, Aemal Akhtar<sup>1,3</sup> , Katie Dawson<sup>1</sup>,  
Phiona Naserian Koyiet<sup>4</sup> and Richard Bryant<sup>1</sup> 

<sup>1</sup>School of Psychology, Science, UNSW Sydney, Sydney, NSW, Australia; <sup>2</sup>School of Psychiatry, Medicine, UNSW Sydney, Sydney, NSW, Australia; <sup>3</sup>Department of Clinical Neuroscience, Division of Insurance Medicine, Karolinska Institutet, Solna, Sweden and <sup>4</sup>Mental Health and Psychosocial Support, World Vision International, Nairobi, Kenya

## Abstract

The current study evaluated the Kiswahili version of General Health Questionnaire (GHQ-12) in a Kenyan context comprising of women exposed to gender-based violence. Participants were randomly drawn from community sampling using household screening methods in peri-urban areas in Nairobi. A total of 1,394 participants with varying levels of literacy (years of education: mean [M] = 9.42; standard deviation [SD] = 3.73) and aged between 18 and 89 years were recruited for the study. The observed factor structure of the GHQ-12 was evaluated using six most tested models querying the dimensionality of the instrument insofar as the impacts of positive and negative wording effects in driving multidimensionality. Results from the confirmatory factor analysis supported a bifactor model, consisting of a general distress factor and two separate factors representing common variance due to the positive and negative wording of items. Overall, the findings support the use of the Kiswahili version of the GHQ-12 as a unidimensional construct with method-specific variance owing to wording effects. Importantly, GHQ-12 responses from a sample of Kenyan women with relatively low levels of literacy are congruent with the factor structure observed in other cross-cultural settings in low- and middle-income countries.

## Impact statement

The highest prevalence rates of common mental disorders are found in low- and middle-income country settings. The need to have efficacious mental health screening tools to detect psychological problems in large numbers of people is important to address mental health needs. As part of a study looking at the impact of a peer-delivered psychological intervention for Kenyan women exposed to gender-based violence, we used the General Health Questionnaire (GHQ-12) to detect general psychological distress. As the factor structure of the Kiswahili version of the GHQ-12 has yet to be validated in this population, we conducted a series of confirmatory factor analyses to elucidate the underlying factor structure. The findings support the proposition that the GHQ-12 can be interpreted as a unidimensional construct with observed factor variation likely arising from wording effects of the items. Importantly, the current findings support and further extend the previous validation completed in a Kenyan context.

## Introduction

Increased exposure to humanitarian crises worldwide has left many people at a heightened risk of developing common mental disorders (CMDs), including depression, anxiety and post-traumatic stress disorder. Individuals with CMDs suffer from wide-ranging physical health problems in ways that contribute to considerable functional impairment and disability (Whiteford et al., 2013; Patel et al., 2016). The heightened prevalence of CMDs among health problems in low- and middle-income countries (LMICs) (GBD, 2022) is not surprising given that relative allocation of global health resources for mental health toward LMICs pales in comparison to high-income countries (Liese et al., 2019). There are apparent challenges in addressing this burden when varied cultural expressions of CMDs and lack of locally validated screening tools impede rapid identification of individuals in need of care and in turn access to appropriate services. For example, CMDs, including major depression, account for the highest burden of disease in Kenya, a trend commonly evidenced across sub-Saharan Africa (Murray et al., 2012).

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

 Cambridge  
Prisms

 CAMBRIDGE  
UNIVERSITY PRESS

Yet, when screening tools developed within Western settings are readily used in culturally diverse settings (without prior local validation), the extent of detection bias in observed prevalence rates is unclear. As such, there is a need to use locally validated screening tools to document the prevalence of CMDs more accurately, inform intervention developments, and, in turn, more effectively allocate limited mental health resources in LMICs such as Kenya.

The first step in this endeavor has often involved the translation of screening tools developed in high-income countries. For example, the 12-item version of the General Health Questionnaire (GHQ-12) is a widely used self-administered screening tool aimed at detecting general psychological distress (Goldberg *et al.*, 1997). The brevity of the GHQ-12 has afforded its administration across clinical and community settings, and this has been done through translation into 38 different languages across several cultural contexts including Africa, Asia, Middle East and Latin America (El-Metwally *et al.*, 2018). Yet, the underlying factor structure of the responses generated from this instrument continues to be debated. The GHQ-12 was originally proposed as a unidimensional measure of psychological distress (Goldberg and Williams, 1988) comprising of equal number of positively and negatively worded items that produce a global distress score. Given the inconsistent evidence for a single-factor structure, alternative two-dimensional and three-dimensional models have been proposed as being more suitable. The two-factor model by Andrich and van Schoubroeck (1989) stipulates factors based on wording comprising of “social dysfunction” (all negatively phrased items) and “anxiety/depression” (all positively worded items). The three-factor model proposed by Graetz (1991) is similar to this with the exception that the negatively phrased items have been split into two distinct factors, namely “loss of confidence” and “anxiety/depression.” This three-factor model by Graetz (1991) has evidenced better fit relative to the one-factor unidimensional model across varying studies in languages other than English (Shevlin and Adamson, 2005; Tomás *et al.*, 2017). Yet, one argument against multidimensionality is that distinct factors formed based on wording of items may represent artifacts of individual response bias instead of theoretically distinct factors. To investigate this hypothesis, wording effects of the GHQ-12 have been modelled in varying ways. First, Hankins (2008) explored a unidimensional model that accounted for wording effects (with two- and three-dimensional models) by correlating error terms of the negatively phrased items to capture systematic error variance, termed a “response bias” model. Second, Ye (2009) extended this approach and modelled an orthogonal (i.e., uncorrelated) method factor associated with negative items, in addition to a general distress factor, termed “method factor” model. Finally, Tomás *et al.* (2017) extended this line of thinking, where two orthogonal method factors are modelled to incorporate negatively and positively phrased items, in addition to the single general distress factor, and termed this as a “bifactor” model. The advantage of this bifactor model has allowed researchers to evaluate common variance among a set of items that can be accounted for by a single unidimensional factor, in addition to variance accounted for by method-specific factors (Reise *et al.*, 2007), not allowed by models proposed by Hankins (2008) and Ye (2009). For example, Centofanti *et al.* (2019) demonstrated the presence of a strong general distress factor as evidenced by the total amount of observed score variance (i.e., omega hierarchical value of 0.81). Yet, the literature is inconsistent as some studies have shown that the multidimensional model by Graetz (1991) outperforms the bifactor model

even after taking wording effects into account (Abubakar and Fischer, 2012; Tomás *et al.*, 2017). The picture is further complicated when findings across studies cannot be easily compared within the context of different scoring methods of the GHQ-12 (Tomás *et al.*, 2017) potentially influencing the observed factor structure and model fit (Rey *et al.*, 2014; Centofanti *et al.*, 2019).

Further, very little is known about the specific factor structure of the Kiswahili (or Swahili) version of this instrument. Investigating the factor structure within this specific context has both practical and theoretical importance. Firstly, Kiswahili is an East African language spoken in more than 14 countries with over 200 million speakers (Lisanza, 2021). As such having the Kiswahili version of the GHQ-12 validated will have wide-ranging practical implications for the selection of locally appropriate screening tools insofar as identification of CMDs in low-resourced sub-Saharan African regions. Second, translation of Western instruments to local languages often brings with it the challenge of finding appropriate wording considered to be equivalent to the Western constructs used in the original instrument. Variations in wording of self-report instruments can often pose unintended consequences in terms of interfering with the subsequent latent structure analysis (Tomás *et al.*, 2013). This is particularly relevant for the context surrounding the use of the Kiswahili version, where psychological distress is commonly expressed through use of local idioms and somatic symptoms (Patel, 1998; Patel and Kleinman, 2003), in turn posing potential variability for the dimensionality of the instrument. The current state of the evidence suggests that the wording of items in the GHQ-12 can indicate differences in response bias or theoretically distinct constructs depending upon the demographic of the sample studied and type of scoring method used. Together, these concerns in the literature motivated the current study to examine the factor structure of Kiswahili version of the GHQ-12 and identify the best fitting model.

The purpose of the present study was to analyze the factor structure of the Kiswahili version of the GHQ-12 administered in a Kenyan context. To the best of our knowledge, one study has previously investigated the factor structure of the GHQ-12 in Kenya (Abubakar and Fischer, 2012); however, this study administered the instrument in a literate English-speaking population of adolescents and adults and did not explore specific reliability estimates (i.e., omega values). We aimed to contribute to the existing debate relating to the dimensionality of the GHQ-12 by choosing models that have garnered the most investigation insofar as ascertaining the extent of wording effects on subsequent observed dimensionality (Abubakar and Fischer, 2012; Hystad and Johnsen, 2020). The current study differs from these previous studies in several ways. We administered the Kiswahili version of the GHQ-12 to a sample of Kenyan women exposed to gender-based violence. Participants were drawn from the community via random household screening in peri-urban Nairobi and thus comprised of varying levels of literacy. This makes our study the first to explore factor structures and associated method effects within an LMIC setting. Here, we evaluate the reliability of the best-fitting model through calculation of omega estimates to better understand the observed factor structure and verify accompanying fit indices, and this, to date, has not been assessed in the target population. To this end, the present study aimed at exploring two research questions: (a) does the Kiswahili version of the GHQ-12 display the same factor structure in a Kenyan population as evidenced in previous studies (Abubakar and Fischer, 2012) and (b) to what extent is the factorial structure of GHQ-12 (in a sample of Kenyan women with varying levels of literacy) influenced by wording effects.

## Methods

### Participants and procedure

The study was carried out in the Dagoretti Sub County, Nairobi, in Kenya. Data were collected across three local health care facilities that are part of the primary health care system. This sample was screened for participation in a larger study on women affected by adversity and gender-based violence in urban areas in Nairobi (Sijbrandij et al., 2016; Bryant et al., 2017). Participants were recruited through random community screening at their households by independent assessors. Households were selected using a population-based interval approach, where members of a larger population were selected according to a random starting point, and a fixed periodic interval. This is called the sampling interval and it is determined by dividing the population size by the desired sample size. Further details on sampling methods have been detailed elsewhere (Sijbrandij et al., 2016). Following selection of households, independent assessors approached the head of each household and asked to interview a random adult woman aged 18 or older. Informed consent was obtained by giving each person 24 h to consider their decision to partake in the trial. For those women who were illiterate, oral consent and a thumb print in lieu of a signature were obtained with an independent witness, in line with WHO recommendations (Bhutta, 2004). Baseline screening for general distress was then administered by an independent assessor at participants' households or a private space near their homes. If a household declined to be screened and/or did not have a woman aged 18 years or older, the independent assessor moved to the next household based on the predetermined sampling interval. Inclusion criteria for this study required that participants were (a) females and (b) over 18 years of age. Data for the GHQ-12 in the current study are based on screening surveys conducted in a total of 1,394 participants.

### Measures

#### GHQ-12

The GHQ-12 is a well-validated indicator of general psychological distress (Goldberg and Williams, 1988) that was used for screening eligible participants in the current study. The GHQ-12 comprises 12 questions about general well-being, experience of depressive and anxiety symptoms and sleep disturbances over “the past few weeks.” In the current study, items were scored on a 4-point Likert scale ranging from 0 to 3 (range 0 to 36; higher scores indicate severe

psychological distress). The GHQ-12 was translated into Kiswahili for the current study and was used to detect psychological comorbidity, as has been previously used in Kenya (Getanda, Papadopoulos, and Evans, 2015).

Prior to screening, this GHQ-12 was translated through a process of cultural adaptation involving two workshops with mental health experts, translators and community health workers. Adaptation of measures focused on cultural appropriateness in terms of language, metaphors, content, concepts and context in line with established guidelines (Bernal and Sáez-Santiago, 2006).

### Data analysis plan

A series of planned confirmatory factor models were examined using MPLUS8 (Muthén and Muthén, 1998–2018) and R packages, including lavaan and psych (Rosseel, 2012). A priori models were estimated using the maximum likelihood estimation (Bentler and Chou, 1987), where this method has been previously used for the 4-point Likert scoring system of the GHQ-12 (King et al., 2023). Other previous studies have also used the variance-adjusted weighted least squares (WLSMV) (Flora and Curran, 2004; Wirth and Edwards, 2007) for ordered categorical items; we carried out analyses with this estimation method as a check for sensitivity, and these WLSMV estimates are available in the [Supplementary material](#). First, our baseline model consisting of a unidimensional model with a single factor as originally stipulated was assessed (see [Figure 1](#)). To explore whether the observed responses from the Kiswahili version of the GHQ-12 represented distinct theoretical factors, models 2 and 3 were chosen (see [Figures 2 and 3](#)). Specifically, a two-factor model as proposed by Andrich and van Schoubroeck (1989) was assessed where one factor included all negatively-worded items and the second factor encompassing all positively worded items. Additionally, a three-factor model proposed by Graetz (1991) was tested. This model was similar to the two-factor model, with the exception that the negatively phrased items were split into two distinct factors: (1) loss of confidence and (2) anxiety and depression. To explore the role of wording effects and related response biases, we chose models 4 to 6. Specifically, model 4 investigated a unidimensional model where errors of all the negatively worded items were correlated as proposed by Hankins (2008) (see [Figure 4](#)). Moreover, model 5 explored a unidimensional model with an orthogonal (i.e., uncorrelated) method factor for the negatively worded items, as proposed by Ye (2009) (see [Figure 5](#)). Finally, model 6 is an extension of model 5, whereby an

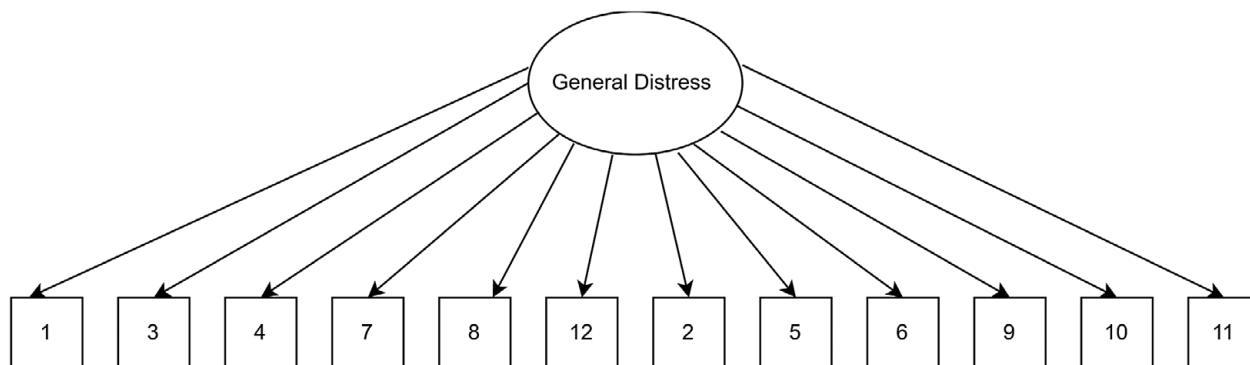


Figure 1. Unidimensional model.

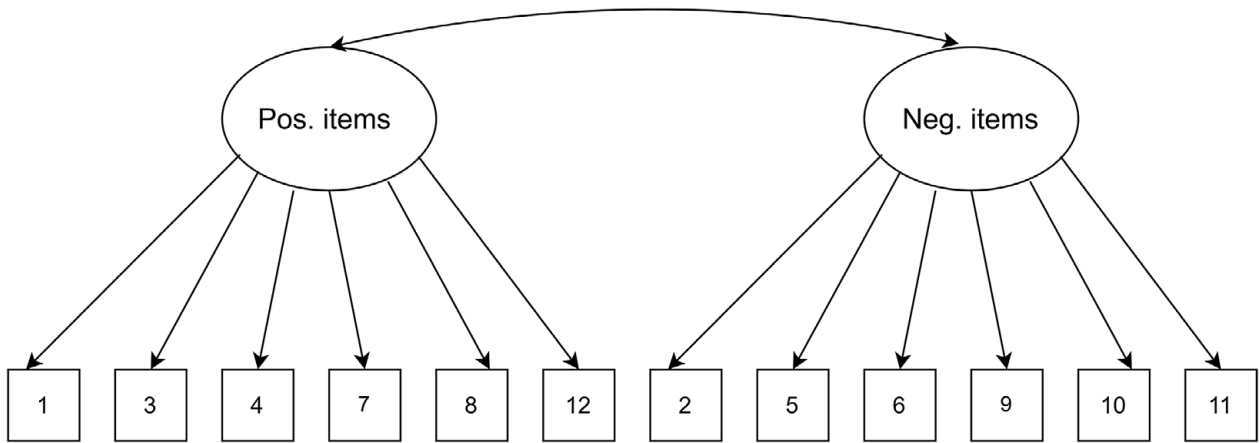


Figure 2. Two-dimensional model.

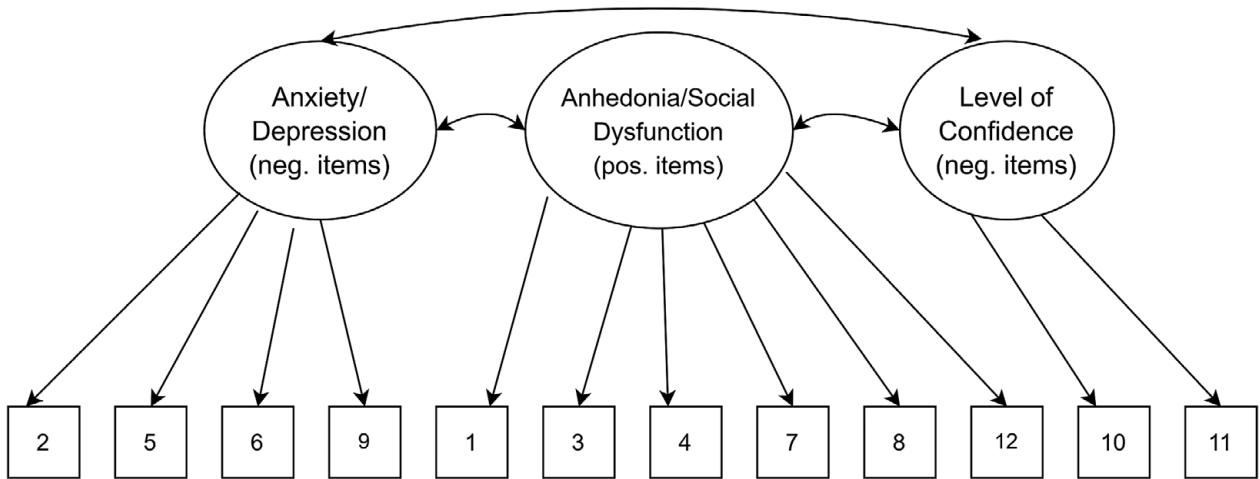


Figure 3. Three-dimensional model.

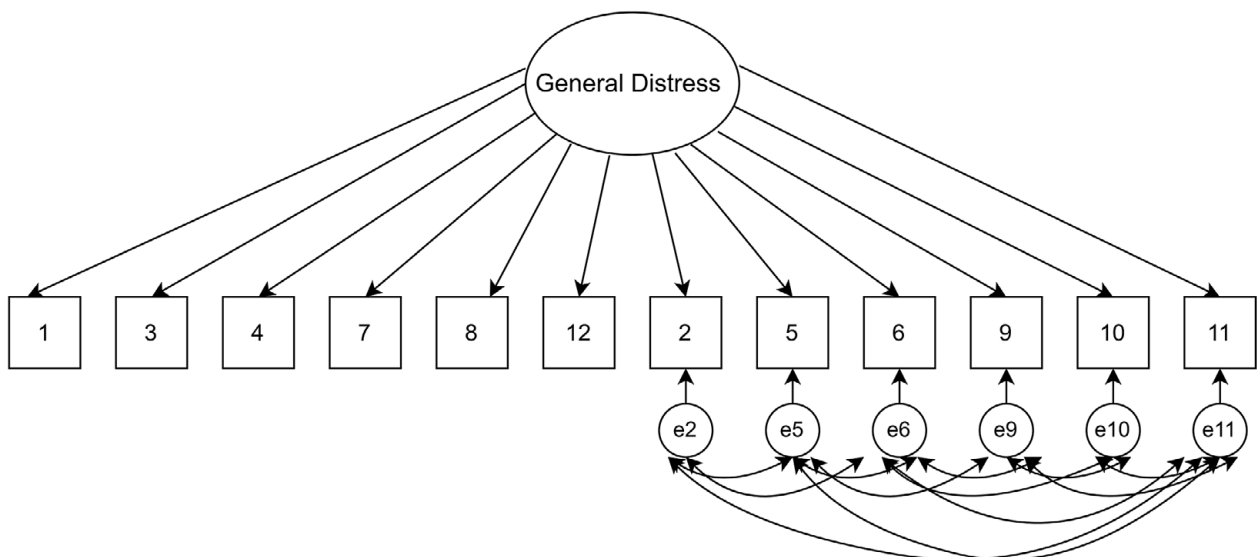


Figure 4. Response bias model.

orthogonal factor for positively phrased items was included (see Figure 6). Also referred to as a bifactor model, this model has been previously tested (Tomás et al., 2017). This bifactor model was chosen due to its validation with the GHQ-12 in recent times (Centofanti et al., 2019) and thus provides an extension of previous validation studies in a Kenyan context (Abubakar and Fischer, 2012).

Individual model fit will be evaluated by examining a combination of the size and statistical significance of factor loadings as well as several commonly used goodness-of-fit statistical indices suggested in the extant literature. Specifically, various parameters

including Comparative Fit Index (CFI), root mean square error of approximation (RMSEA) and Tucker–Lewis index (TLI) were chosen. Good model fit has been shown to be reflected in CFI and TLI values greater than or equal to 0.95 (Hu and Bentler, 1999). Additionally, RMSEA values reflect sensitivity to model misspecification where values closer to 0.06 have been proposed to indicate good model fit. As the chi-square metric is known to be sensitive to sample size, it was used in combination with the above fit indices (Schumacker and Lomax, 2004). Additionally, comparative model fit will be examined with two measures including the Akaike information criterion and the Bayesian information

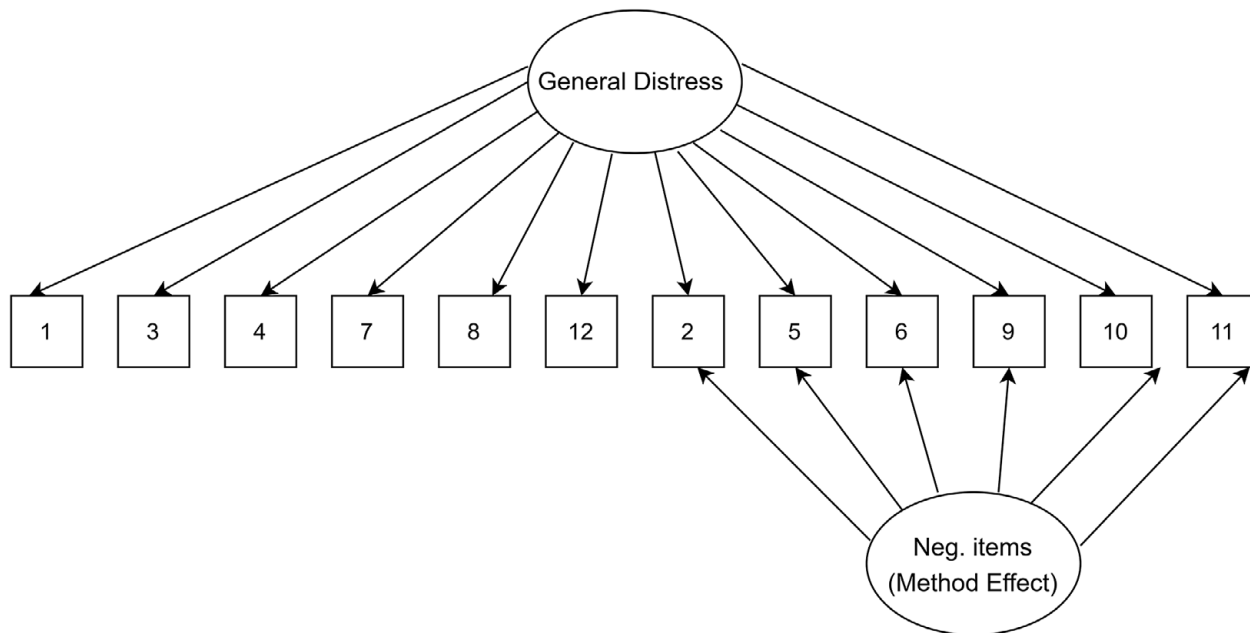


Figure 5. Method factor model.

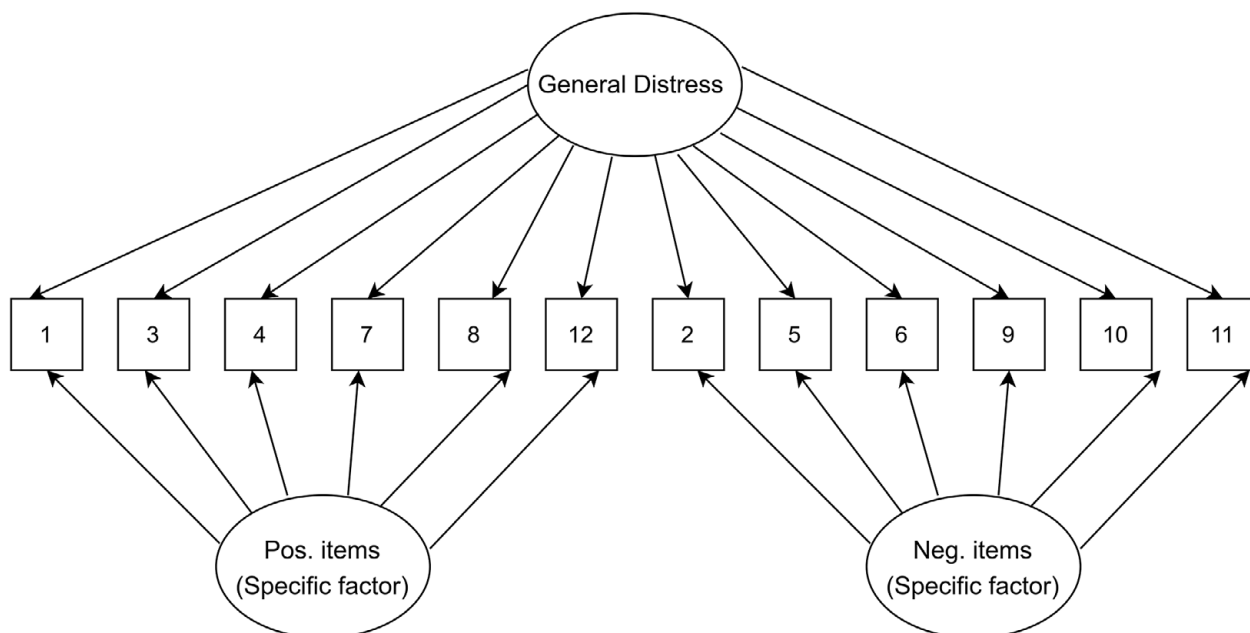


Figure 6. Bifactor model.

criterion, where lower values for both measures indicate better fit. Finally, to explore the reliability of observed dimensionality, we computed omega estimates including omega hierarchical ( $\omega_h$ , proportion of total score variance owing to general factor over and above specific factors) and omega total ( $\omega$ , proportion of total variance across general and specific factors) scores (McNeish, 2018).

## Results

### Demographic characteristics

Data were collected from 1,394 adult females aged 18 years and above, with a mean age of 32.82 years (standard deviation [SD] = 11.539, range = 18–89). Years of education ranged from 0 to 24 years (mean [M] = 9.42, SD = 3.73). Specifically, 11% of women had five or less years of education and 60% had 10 or less years of education.

### Confirmatory factor analyses

Table 1 presents fit indices of the models assessed. The single-factor unidimensional model (Figure 1) was initially tested, and fit indices indicated that this model provided the poorest fit of all the models

tests (CFI = 0.902; TLI = 0.878; standardized root mean square residual [SRMR] = 0.050; RMSEA = 0.093 with 90% confidence interval [CI] for RMSEA = 0.087 - 0.099). The two- (Figure 2) and three-dimensional (Figure 3) models (without method effects) improved these fit statistics, where the latter provided more acceptable fit on the RMSEA (RMSEA = 0.046 with 90% CI for RMSEA = 0.040–0.053). After adjusting for correlated errors between items, the response bias model (Figure 4) had a similar fit to the three-dimensional model (RMSEA = 0.047 with 90% CI for RMSEA = 0.040–0.055). The method factor model (Figure 5) with the artifactual factor containing all the negative items did not fit the data better than the three-dimensional model (CFI = 0.953; TLI = 0.935; SRMR = 0.034; RMSEA = 0.068 with 90% CI for RMSEA = 0.061–0.075). The bifactor model (Figure 6) provided good fit statistics on all indices (See Table 1). The standardized factor loadings for the bifactor model are presented in Table 3, and for all other models in Table 2. All absolute residual correlations were small for the bifactor model (<0.04).

To explore the amount of observed score variance accounted for by the general factor relative to specific factors of the bifactor model, different forms of omega estimates were computed. First, the omega ( $\omega$ ) value is a reliability estimate of a unit-weighted total score of all GHQ items representing both the general and specific

**Table 1.** Model fit indices

Model	$\chi^2$	df	p value	AIC	BIC	CFI	TLI	RMSEA (90% CI)	SRMR
1. Unidimensional	682.179	53	0.000	35,419.19	35,612.43	0.902	0.878	0.093 (0.087–0.099)	0.050
2. Two-dimensional	408.204	53	0.000	34,942.10	35,145.78	0.945	0.931	0.070 (0.064–0.076)	0.037
3. Three-dimensional	201.089	51	0.000	34,942.10	35,145.78	0.977	0.970	0.046 (0.040–0.053)	0.028
4. Response bias (correlated errors)	158.317	39	0.000	34,923.33	35,189.68	0.981	0.969	0.047 (0.040–0.055)	0.026
5. Method factor	350.211	48	0.000	35,097.22	35,316.57	0.953	0.935	0.068 (0.061–0.075)	0.034
6. Bifactor	132.940	42	0.000	34,891.95	35,142.64	0.986	0.978	0.040 (0.032–0.047)	0.020

AIC, Akaike information criterion; BIC, Bayesian information criterion; CI, confidence interval; CFI, Comparative Fit Index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; TLI, Tucker–Lewis index.

**Table 2.** Standardized factor loadings for the alternative models of the GHQ-12

Items	Model 1		Model 2		Model 3			Model 4	Model 5		Model 6		
	I	II	I	II	I	II	III	I	I	II	I	II	III
1 Able to concentrate	0.572	0.628			0.626			0.624	–0.620		0.518	0.340	
2 Loss of sleep over worry	0.698		0.718	0.724				0.558	–0.563	0.461	0.704		0.136
3 Playing a useful part	0.480	0.537			0.536			0.535	–0.530		0.419	0.394	
4 Capable of making decisions	0.378	0.457			0.458			0.458	–0.454		0.308	0.453	
5 Felt constantly under strain	0.757		0.781	0.802				0.612	–0.606	0.536	0.771		0.290
6 Could not overcome difficulties	0.754		0.765	0.774				0.649	–0.643	0.413	0.750		0.193
7 Able to enjoy day-to-day activities	0.606	0.683			0.684			0.686	–0.682		0.549	0.407	
8 Able to face problems	0.627	0.666			0.664			0.664	–0.658		0.581	0.273	
9 Feeling unhappy and distressed	0.768		0.789	0.799				0.629	–0.632	0.487	0.776		0.162
10 Losing confidence	0.637		0.633			0.755	0.577	–0.609	0.182	0.695			–0.397
11 Thinking of self as worthless	0.672		0.674			0.800	0.582	–0.618	0.246	0.715			–0.267
12 Feeling reasonably happy	0.598	0.653			0.656			0.656	–0.658		0.554	0.321	

**Table 3.** Standardized factor loadings and variance composition for the bifactor model of the General Health Questionnaire (GHQ-12)

	Model 6: Bifactor Model		
	I	II	III
Able to concentrate	0.518	0.340	
Loss of sleep over worry	0.704		0.136
Playing a useful part	0.419	0.394	
Capable of making decisions	0.308	0.453	
Felt constantly under strain	0.771		0.290
Could not overcome difficulties	0.750		0.193
Able to enjoy day-to-day activities	0.549	0.407	
Able to face problems	0.581	0.273	
Feeling unhappy and distressed	0.776		0.162
Losing confidence	0.695		-0.397
Thinking of self as worthless	0.715		-0.267
Feeling reasonably happy	0.554	0.321	
ECV	0.850	0.073	0.078
$\Omega$	0.906		
$\omega_s$		0.783	0.892
$\omega_h$	0.832		
$\omega_{hs}$		0.281	0.001

ECV, explained common variance;  $\omega$ , omega;  $\omega_s$ , omega subscale;  $\omega_h$ , omega hierarchical;  $\omega_{hs}$ , omega hierarchical subscale.

factors and is based on the sums of squared loadings and error variances. Here, approximately 91% of variance in this total scale score is accounted for by the combination of the general and specific factors ( $\omega = 0.906$ ). Specifically, approximately 83% of the variance in the observed scale score is attributed to just the general factor (omega hierarchical;  $\omega_h = 0.832$ ). After controlling for the general factor, approximately 28% of the variance in the positive subscale score is accounted for by the specific factor (Omega hierarchical subscale;  $\omega_{hs} = 0.281$ ), and approximately 0.1% variance accounted for by the negative subscale score for the remaining specific factor ( $\omega_{hs} = 0.001$ ). As seen in Table 3, it is worth noting that omega hierarchical values ( $\omega_{hs}$ ) were very low relative to the omega subscale scores ( $\omega_s$ ), suggesting that a large proportion of the variance in subscale scores can be attributed to the general factor as opposed to unique specific factor variances.

## Discussion

This study examined the factorial structure of the Kiswahili version of the GHQ-12 in a sample of Kenyan women exposed to gender-based violence. To our knowledge, this is the first study to assess the five alternative models proposed in the literature. Our findings indicate that the GHQ-12 factor structure as derived from a sample of Kenyan women with varying levels of literacy displays a general distress factor with two separate factors resulting from method-specific wording effects. This is similar to recent validation of the GHQ-12 within other demographic samples comprising of literate English-speaking individuals (Centofanti et al., 2019; Hystad and Johnsen, 2020). Importantly, the current findings support the

proposition that variation in factors across multidimensional models (i.e., two- and three-factor models) may be largely related to differential phrasing of items. Specifically, the finding that negatively and positively worded items load onto separate factors is consistent with suggestions of method-specific variance in the GHQ-12 that likely drives its observed multidimensionality (Hankins, 2008; Ye, 2009). The current study extends previous support for the unidimensional GHQ-12 in an English-speaking literate Kenyan population (Abubakar and Fischer, 2012) by considering both positive and negative wording effects (Marsh et al., 2010). In the current analysis, the general factor accounted for nearly 91% of the total score variance ( $\omega = 0.906$ ), while the specific method factors accounted for very little in comparison, after controlling for the general factor variance ( $\omega_s = 0.001-0.281$ ). It is worth noting that the three-dimensional model displayed a good fit that was comparable to the bifactor model, suggesting that the plausibility of the three-dimensional model cannot be precluded based on observed fit indices alone. However, the practical utility in having three factors is questioned within the context of high correlations among the three factors and high standardized factor loadings for the latent factors. Here, our sensitivity analyses using the WLSMV estimator evidence similar conclusions regarding the bifactor and three-factor models (see Supplementary material). To this end, previous research has questioned the discriminative power among the three factors and suggested that the three factors do not afford additional predictive power (e.g., Ye, 2009). Further support for this is seen in the goodness-of-fit indices of the two- and three-dimensional models that fell within close range of the unidimensional models fitted with method effects (i.e., models 4 and 5), in turn suggesting little incremental value for having theoretical distinct factors. Importantly, we followed up the observation that the bifactor model may offer the most sufficient explanation of the GHQ-12 through computation of specific omega estimates. Taken together, we find that the bifactor model comprising of a single general distress factor and two separate factors capturing method-specific variance offers the most parsimonious explanation of the GHQ-12 factor structure. Here, it is worth considering that summation of items on the GHQ-12, as commonly scored, may reflect the influence of item phrasing that is beyond the dominant general distress factor.

Nonetheless, the current findings offer important practical relevance for administering the GHQ-12 in cross-cultural settings. From an applied perspective, the findings suggest that the Kiswahili version of the GHQ-12 may be administered in a Kenyan population and interpreted in a similar manner to the English version of this measure administered in a demographical similar population (Abubakar and Fischer, 2012). Here, screening for general distress including symptoms of CMDs including anxiety and depression as measured by the GHQ-12 in this setting offers similar clustering as compared to other cultural settings around the world (Tomás et al., 2017). This is an important reflection as some research has previously suggested that clustering of items in the GHQ-12 item may differ across cultural contexts (Romppel et al., 2017). However, this proposed functional equivalence of the GHQ-12 requires further statistical verification insofar as performing tests of measurement invariance. Specifically, comparison of the GHQ-12 across literacy groups and across time will provide important information on the predictive utility of the observed response patterns.

To this end, we acknowledge several limitations to the current study. First, the generalizability of our findings needs to be assessed further to mix-gendered samples in the same population. Second, the current analysis was limited to the 4-point Likert rating method

of the GHQ-12; here, scoring system can influence the number of factors obtained as well as item-factor loadings (Aguado et al., 2012) with other studies using different scoring systems, including 6-point (Kalliath, O'Driscoll, and Brough, 2004) and 7-point (Ye, 2009) scales. Third, we recognize that this sample was challenged in levels of literacy as measured by years of education, and this may have influenced response patterns insofar as how items are perceived and responded to by the target sample. Fourth, tests of measurement invariance will help further consolidate our proposition that the GHQ-12 can be practically interpreted in ways similar to the English version of this instrument.

These limitations notwithstanding the current findings strengthen support for the proposition that the GHQ-12 is a unidimensional measure with factorial variation that likely arises from wording effects. Our findings are strengthened by reports of omega reliability estimates and this is an extension of the research previously conducted in Kenya (Abubakar and Fischer, 2012). From a practical or applied perspective, it may be acceptable to utilize this instrument as a unidimensional measure within the cultural context of Kenya.

**Open peer review.** To view the open peer review materials for this article, please visit <http://doi.org/10.1017/gmh.2024.46>.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/gmh.2024.46>.

**Data availability statement.** Data and materials will be available on request to the corresponding author, D.K.

**Acknowledgements.** The authors alone are responsible for the views expressed in this article, which do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

**Author contribution.** All authors are responsible for the study concept and design. D.K. and D.H-P. developed the methods and conducted the data analysis. D.K. wrote the first draft of the manuscript, and D.H-P., A.A., and R.B. contributed to the critical revision of the final manuscript. All authors reviewed and approved the final manuscript.

**Financial support.** This study was supported by Grand Challenges Canada #0368-04, World Vision Canada and World Vision Australia.

**Competing interest.** The authors declare none.

**Ethics statement.** The data from this study were drawn from a clinical trial approved by the WHO Research Ethics Review Committee (RPC656) and the Great Lakes University Ethics Committee in Kenya.

## References

- Abubakar A and Fischer R (2012) The factor structure of the 12-item general health questionnaire in a literate Kenyan population. *Stress and Health* **28**(3), 248–254. <https://doi.org/10.1002/smi.1420>.
- Aguado J, Campbell A, Ascaso C, Navarro P, García-Estevé L and Luciano JV (2012) Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish postpartum women. *Assessment* **19**, 517–525.
- Andrich D and van Schoubroeck L (1989) The general health questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine* **19**(2), 469–485. <https://doi.org/10.1017/S0033291700012502>.
- Bentler PM and Chou C-P (1987) Practical issues in structural modeling. *Sociological Methods & Research* **16**(1), 78–117. <https://doi.org/10.1177/0049124187016001004>.
- Bernal G and Sáez-Santiago E (2006) Culturally centered psychosocial interventions. *Journal of Community Psychology* **34**(2), 121–132. <https://doi.org/10.1002/jcop.20096>.
- Bhutta ZA (2004) Beyond informed consent. *Bulletin of the World Health Organization* **82**(10), 771–777.
- Bryant RA, Schafer A, Dawson KS, Anjuri D, Mulili C, Ndogoni L, Koyiet P, Sijbrandij M, Ulate J, Harper Shehadeh M, Hadzi-Pavlovic D and van Ommeren M (2017) Effectiveness of a brief behavioural intervention on psychological distress among women with a history of gender-based violence in urban Kenya: A randomised clinical trial. *PLoS Medicine* **14**(8), e1002371. <https://doi.org/10.1371/journal.pmed.1002371>.
- Centofanti S, Lushington K, Wicking A, Wicking P, Fuller A, Janz P and Dorrian J (2019) Establishing norms for mental well-being in young people (7–19 years) using the general health questionnaire-12. *Australian Journal of Psychology* **71**(2), 117–126. <https://doi.org/10.1111/ajpy.12227>.
- El-Metwally A, Javed S, Razzak HA, Aldossari KK, Aldiab A, Al-Ghamdi SH ... Al-Zahrani JM (2018) The factor structure of the general health questionnaire (GHQ12) in Saudi Arabia. *BMC Health Services Research* **18**(1), 595. <https://doi.org/10.1186/s12913-018-3381-6>.
- Flora DB and Curran PJ (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods* **9**(4), 466–491. <https://doi.org/10.1037/1082-989X.9.4.466>.
- GBD (2022) Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *The Lancet Psychiatry* **9**(2), 137–150. [https://doi.org/10.1016/S2215-0366\(21\)00395-3](https://doi.org/10.1016/S2215-0366(21)00395-3).
- Getanda EM, Papadopoulos C and Evans H (2015) The mental health, quality of life and life satisfaction of internally displaced persons living in Nakuru County, Kenya. *BMC public health* **15**, 755. <https://doi.org/10.1186/s12889-015-2085-7>.
- Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O and Rutter C (1997) The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine* **27**(1), 191–197. <https://doi.org/10.1017/S0033291796004242>.
- Goldberg DP and Williams P (1988) *A User's Guide to the General Health Questionnaire*. Windsor, UK: NFER-Nelson.
- Graetz B (1991) Psychiatric epidemiology multidimensional properties of the general health questionnaire. *Social Psychiatry and Psychiatric Epidemiology* **26**(3), 132–138.
- Hankins M (2008) The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC Public Health* **8**, 1–7. <https://doi.org/10.1186/1471-2458-8-355>.
- Hu L-t and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* **6**(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Hystad SW and Johnsen BH (2020) The dimensionality of the 12-item general health questionnaire (GHQ-12): Comparisons of factor structures and invariance across samples and time [original research]. *Frontiers in Psychology* **11**, 1300. <https://doi.org/10.3389/fpsyg.2020.01300>.
- Kalliath TJ, O'Driscoll MP and Brough P (2004) A confirmatory factor analysis of the General Health Questionnaire-12. *Stress and Health: Journal of the International Society for the Investigation of Stress* **20**(1), 11–20.
- King K, Allum N, Stoneman P and Cernat A (2023) Estimating measurement equivalence of the 12-item general health questionnaire across ethnic groups in the UK. *Psychological Medicine* **53**(5), 1778–1786. <https://doi.org/10.1017/S0033291721003408>.
- Liese BH, Gribble RSF and Wickremesinhe MN (2019) International funding for mental health: A review of the last decade. *International Health* **11**(5), 361–369. <https://doi.org/10.1093/inthealth/ihz040>.
- Lisanza V (2021) Swahili Gaining Popularity Globally. Available at <https://www.un.org/africarenewal/magazine/december-2021/swahili-gaining-popularity-globally> (accessed 3 July 2021).
- Marsh HW, Scalas LF and Nagengast B (2010) Longitudinal tests of competing factor structures for the Rosenberg self-esteem scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment* **22**(2), 366–381. <https://doi.org/10.1037/a0019225>.
- McNeish D (2018) Thanks coefficient alpha, we'll take it from here. *Psychological Methods* **23**(3), 412–433. <https://doi.org/10.1037/met0000144>.
- Muthén LK and Muthén BO (1998–2018) *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.



- Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, ... Memish ZA** (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *Lancet* **380**(9859), 2197–2223. [https://doi.org/10.1016/s0140-6736\(12\)61689-4](https://doi.org/10.1016/s0140-6736(12)61689-4).
- Patel V** (1998) *Culture and Common Mental Disorders in Sub-Saharan Africa*. East Sussex, UK: Psychology Press.
- Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, ... Whiteford H** (2016) Addressing the burden of mental, neurological, and substance use disorders: Key messages from disease control priorities, 3rd Edn. *The Lancet* **387**(10028), 1672–1685. [https://doi.org/10.1016/S0140-6736\(15\)00390-6](https://doi.org/10.1016/S0140-6736(15)00390-6).
- Patel V and Kleinman A** (2003) Poverty and common mental disorders in developing countries. *Bulletin of the World Health Organization* **81**(8), 609–615.
- Reise SP, Morizot J and Hays RD** (2007) The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research* **16**(SUPPL. 1), 19–31. <https://doi.org/10.1007/s11136-007-9183-7>.
- Rey JJ, Abad FJ, Barrada JR, Garrido LE and Ponsoda V** (2014) The impact of ambiguous response categories on the factor structure of the GHQ–12. *Psychological Assessment* **26**, 1021–1030.
- Romppel M, Hinz A, Finck C, Young J, Brähler E and Glaesmer H** (2017) Cross-cultural measurement invariance of the general health Questionnaire-12 in a German and a Colombian population sample. *International Journal of Methods in Psychiatric Research* **26**(4), e1532 <https://doi.org/10.1002/mpr.1532>.
- Rosseel Y** (2012) Lavaan: An R package for structural equation modeling. *Journal of Statistical Software* **48**(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Schumacker RE and Lomax RG** (2004) *A beginner's guide to structural equation modeling* (2nd ed.). Lawrence Erlbaum Associates Publishers.
- Shevlin M and Adamson G** (2005) Alternative factor models and factorial invariance of the GHQ-12: A large sample analysis using confirmatory factor analysis. *Psychological Assessment* **17**(2), 231–236. <https://doi.org/10.1037/1040-3590.17.2.231>.
- Sijbrandij M, Bryant RA, Schafer A, Dawson KS, Anjuri D, Ndogoni L, ... van Ommeren M** (2016) Problem management plus (PM+) in the treatment of common mental disorders in women affected by gender-based violence and urban adversity in Kenya; Study protocol for a randomized controlled trial. *International Journal of Mental Health Systems* **10**, 44. <https://doi.org/10.1186/s13033-016-0075-5>.
- Tomás JM, Gutiérrez M and Sancho P** (2017) Factorial validity of the general health questionnaire 12 in an Angolan sample. *European Journal of Psychological Assessment* **33**(2), 116–122. <https://doi.org/10.1027/1015-5759/a000278>.
- Tomás JM, Oliver A, Galiana L, Sancho P and Lila M** (2013) Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling* **20**(2), 299–313. <https://doi.org/10.1080/10705511.2013.769394>.
- Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, ... Vos T** (2013) Global burden of disease attributable to mental and substance use disorders: Findings from the global burden of disease study 2010. *Lancet* **382**(9904), 1575–1586. [https://doi.org/10.1016/s0140-6736\(13\)61611-6](https://doi.org/10.1016/s0140-6736(13)61611-6).
- Wirth RJ and Edwards MC** (2007) Item factor analysis: current approaches and future directions. *Psychological methods* **12**(1), 58–79. <https://doi.org/10.1037/1082-989X.12.1.58>.
- Ye S** (2009) Factor structure of the general health questionnaire (GHQ-12): the role of wording effects. *Personality and Individual Differences* **46**(2), 197–201. <https://doi.org/10.1016/j.paid.2008.09.027>.