

# Are Sinhala and Tamil PHQ-9 versions comparable across the Sri Lankan population? Psychometric equivalence of the SLHAS PHQ-9 form across translations and between demographic subgroups

Sasanka<sup>1</sup>, Ukwaththage Bhimal Eshan<sup>1</sup>, Kumar<sup>1</sup>, Shamala Rannan-Eliya<sup>2</sup>, Ravindra Prasan<sup>1</sup>

(Key words: depression, mental health, PHQ-9, psychometric equivalence, Sri Lanka)

## Abstract

**Introduction:** Prevalence estimates of depressive disorder in Sri Lanka vary widely, partly due to insufficient validation of screening tools in the Sri Lankan population and across major demographic subgroups.

**Objective:** This study evaluates the validity and psychometric equivalence of the Patient Health Questionnaire-9 (PHQ-9), a widely used measure of depression, across Sinhala and Tamil language translations, and different demographic groups.

**Methods:** Using data from a nationally representative sample of 6,444 participants in the 2018-19 Wave 1 of the Sri Lanka Health and Ageing Study (SLHAS), we assessed the PHQ-9's factor structure, convergent and divergent validity, and measurement invariance. Multigroup confirmatory factor analysis was employed to test for equivalence across translations, gender, ethnicity, education, and socioeconomic status.

**Results:** Results supported a two-factor model of the PHQ-9 and demonstrated evidence of convergent and divergent validity. They also indicated equivalence across the two translated versions, and across gender, ethnicity, level of education, and socioeconomic status.

**Conclusion:** The SLHAS version of the PHQ-9 is a valid and reliable tool for screening of depressive symptoms and for conducting research in Sri Lanka. Its equivalence across translations and diverse demographic characteristics ensures its suitability for use in this population.

## Introduction

Depressive disorders make a major contribution to global morbidity and impose a significant social burden,

and the available evidence suggests that prevalence in Sri Lanka may exceed the regional average. Despite this, Sri Lanka lacks consistent and robust estimates of prevalence and trends. Existing studies reveal considerable heterogeneity in prevalence estimates across studies and different demographic groups. The lack of comparability in the screening tools that different studies have used to assess depressive disorders, reliance on small or non-representative population samples, and limited evidence of the performance and validity of established survey instruments across different Sri Lankan population groups may account for inconsistencies and heterogeneity [1].

The most comprehensive estimate of mental health disorders in Sri Lanka is by the National Mental Health Survey (NMHS) in 2007 [2], but it was limited to adults aged 65 years and below and excluded the Northern and Eastern provinces due to the then ongoing war. Using the Patient Health Questionnaire with 9 items (PHQ-9) to assess levels of depressive symptoms, the NMHS estimated that 10% of the population experienced mental disorders, with 6.9% and 2.4% reporting mild to moderate depression and major depression, respectively. Although the study instruments were administered in both Sinhala and Tamil, the NMHS did not discuss or consider how its PHQ-9 instrument performed across its Sinhala and Tamil translations, or across different population subgroups.

Variability in instrument performance across groups can distort comparisons, particularly when scores are compared or pooled (see for instances [3]), as score differences may reflect methodological inconsistencies rather than true prevalence differences. Factors influencing prevalence estimates include differing operational

*Ceylon Medical Journal* 2023; **68**: 174-183

DOI: <https://10.4038/cmj.v68i4.9946>

<sup>1</sup>Department of Agricultural Economics and Business Management, Faculty of Agriculture, University of Peradeniya, Sri Lanka, <sup>2</sup>Institute for Health Policy, Sri Lanka.

Correspondence: SK, e-mail: <[skumar@agri.pdn.ac.lk](mailto:skumar@agri.pdn.ac.lk)>. Received 29 May 2023 and revised version 25 November 2023 accepted 11 December 2023



This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

definitions of ‘depression’, severity thresholds, and psychometric characteristics of the tools used. Reliable prevalence measurement also requires instruments validated across languages and population subgroups [4]. Measurement invariance testing, using multiple-group confirmatory factor analysis (CFA), can assess an instrument’s psychometric performance across groups and whether the same underlying factor structure can be assumed to exist across these groups [5]. Only when invariance is established can differences in scores across groups be assumed to represent actual group differences in depression [6,7]. To our knowledge, research in Sri Lanka has largely neglected the issue of scale equivalence across local languages and population subgroups [see for instance, 8-13].

### Patient Health Questionnaire- 9 (PHQ-9) and its factor structure

The Patient Health Questionnaire-9 (PHQ-9) [14] is a widely used tool for assessing depressive symptoms, including population prevalence [14-18]. While the PHQ-9 has been used in Sri Lanka [10,12,19-21], we were unable to find previous studies in the literature examining its factor structure and measurement invariance across the two major local languages or key socio-demographic groups in Sri Lanka. The only documented validation of the PHQ-9 in Sri Lanka involved a Sinhala translation in a study of a sample of 75 clinically diagnosed depression patients, from an outpatient psychiatry clinic in a tertiary care hospital in Colombo, and 75 gender-matched controls from the community, which demonstrated acceptable sensitivity and specificity [22].

In the global literature, two conceptualizations of the PHQ-9’s factor structure [15] have been proposed. The first supports a single factor model [23-28], while the second posits a two factor model distinguishing cognitive/affective (non-somatic) and somatic dimensions [15,23-25,27-31]. Debate exists over item classification within the two factor model [15]. The most accepted version [15,29,31] classifies three somatic items (sleep disturbance, fatigue, appetite changes), and six cognitive/affective items. Others include psychomotor difficulties or concentration problems within the somatic dimension [15,29,30] forming variations of the two subscales [15,31].

This study aimed to evaluate the construct validity of the PHQ-9 instrument used by the Sri Lanka Health and Ageing Study (SLHAS) in the Sri Lankan populations by: (1) assessing its factor structure; (2) testing measurement invariance across Sinhala and Tamil translations and demographic groups (gender, ethnicity, education, and socioeconomic status); and (3) examining its convergent and divergent validity.

## Methods

### Data source, sample, and measures

Our main analysis uses data from the Wave 1 (W1) of the SLHAS, conducted from mid-November 2018 to mid-November 2019, with a sample of 6,627 participants (see Table 1 for sample details). Further information about the SLHAS is available in previous SLHAS publications [32-34]. A key objective of the SLHAS is to track changes in mental health in the population, so reliability of its mental health instruments is a general concern.

Obtaining validated versions of the PHQ-9 in Sinhala and Tamil for use in the SLHAS proved difficult. SLHAS investigators were unable to find a Tamil version, and the previously reported Sinhala version [22] could not be obtained despite numerous efforts to contact the corresponding author of the relevant study. Three items, 6, 7 and 8 were simplified to read “*Feeling bad about yourself or that you are a failure*”, “*Trouble concentrating on things*” and “*Moving or speaking so slowly that other people could have noticed*”, for which the original versions read as “*Feeling bad about yourself or that you are a failure or have let yourself or your family down*”, “*Trouble concentrating on things, such as reading the newspaper or watching television*” and “*Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual*”, respectively. The English version of the PHQ-9 was then translated into Sinhala and Tamil (see Supplemental materials). This was done through a process of translation and back translation, both of which were conducted by a team of researchers and language experts. The final translations were implemented in the survey’s Computer-Assisted Personal Interview (CAPI) software, which ran on computer tablets, in the form of separate bilingual Sinhala/English and Tamil/English versions, with the English text being displayed immediately after the Sinhala or Tamil text of each question.

To assess the robustness of our analysis to the wording difference in the PHQ-9, we examined item 8, between the standard version and that used in the SLHAS W1 survey, through the extended second wave (W2x) of the SLHAS, using interviews conducted by telephone during June-September 2023. The W2x survey reinterviewed original members from the W1 cohort plus additional respondents recruited by random digit dialing, and it used the PHQ-8 (8 item) version of the instrument, which excludes the last item (suicidal ideation), to assess depression symptoms [35]. For a subset of W2x respondents, item 8 was randomized to either the truncated Wave 1 wording or the standard full wording to assess potential impacts of the wording difference in the SLHAS W1 instrument.

Table 1. Characteristics of respondent sample

	<i>Frequency (%)</i>	<i>PHQ-9 Total Score Mean (SD*)</i>	<i>PHQ-9 Total Score ≥ 10 (%)</i>	<i>Frequency, who reported as they have depression (%)</i>	<i>Frequency, who has been diagnosed in medical reports as having depression (%)</i>	<i>Frequency, who is taking medication or has been prescribed medication for depression (%)</i>
<b>Language of the questionnaire (n=6,346)</b>						
Sinhala	4,563 (71.9)	3.03 (3.82)	6.8	40 (0.9)	22 (0.5)	27 (0.6)
Tamil	1,783 (28.1)	2.82 (3.59)	5.7	23 (1.3)	11 (0.6)	14 (0.8)
<b>Gender (n=6,346)</b>						
Male	3,106 (48.9)	2.91 (3.74)	6.1	29 (0.9)	19 (0.6)	21 (0.7)
Female	3,240 (51.1)	3.03 (3.77)	6.9	34 (1.0)	14 (0.4)	20 (0.6)
<b>Ethnicity (n=6,346)</b>						
Sinhala	4,466 (70.4)	2.99 (3.80)	6.7	37 (0.8)	18 (0.4)	23 (0.5)
Sri Lankan Tamil	1,238 (19.5)	2.90 (3.63)	6.0	17 (1.3)	10 (0.8)	11 (0.8)
Muslim	268 (4.2)	2.95 (3.76)	7.2	3 (1.1)	1 (0.4)	3 (1.1)
Indian Tamil	242 (3.8)	2.98 (3.72)	5.4	2 (0.8)	2 (0.8)	2 (0.8)
Other	132 (2.1)	2.89 (3.70)	5.8	0 (0.0)	0 (0.0)	0 (0.0)
<b>Education Level (n=6,346)</b>						
No schooling	242 (3.8)	3.19 (3.71)	7.0	4 (1.7)	0 (0.0)	1 (0.4)
Grade 1-5	890 (14.0)	2.80 (3.76)	5.7	9 (1.0)	3 (0.3)	5 (0.6)
Grade 6-12	2,324 (36.6)	2.98 (3.82)	6.9	18 (0.8)	15 (0.6)	13 (0.6)
Passed O/L**	1,563 (24.6)	2.95 (3.68)	6.3	16 (1.0)	7 (0.4)	11 (0.7)
Passed A/L***	1,069 (16.8)	3.05 (3.74)	6.3	11 (1.0)	6 (0.6)	8 (0.7)
Degree & above	258 (4.1)	3.03 (3.84)	6.6	1 (0.4)	0 (0.0)	1 (0.4)
<b>Socio-Economic Status (n=6,346)</b>						
Poorest	1,549 (24.4)	2.84 (3.72)	6.2	16 (1.0)	9 (0.6)	10 (0.6)
Poorer	1,270 (20.0)	2.92 (3.75)	6.7	10 (0.8)	4 (0.3)	7 (0.6)
Middle	1,189 (18.7)	3.09 (3.93)	7.1	18 (1.5)	10 (0.8)	12 (1.0)
Richer	1,129 (17.8)	3.04 (3.66)	6.5	7 (0.6)	3 (0.3)	3 (0.3)
Richest	1,209 (19.1)	3.00 (3.76)	7.8	12 (1.0)	7 (0.6)	9 (0.7)

Note: \*SD = Standard Deviation; \*\*Ordinary Level Examinations; \*\*\*Advanced Level Examinations.

## Data Analysis

Given the ordinal nature of the PHQ-9 scale, a polychoric correlation matrix was used for confirmatory factor analyses (CFA; [36]), using the FACTOR software [37]. The matrix was created with varimax rotation. Kaiser-Meyer-Olkin (KMO) and Bartlett Sphericity tests, yielding a KMO value of 0.91 and a Bartlett's statistic of 26,778.3 (df = 36;  $p < .001$ ), indicated suitability for factor analysis.

To determine the factor structure and establish invariance, a series of CFAs, using unweighted least square (ULS) estimation [38], were performed using SPSS AMOS version 26 with the pre-calculated polychoric matrix. Once the appropriate model was selected through the process described below, invariance was tested.

Four models drawn from the literature were assessed to establish the factor structure of the PHQ-9:

- **Model 1:** Single factor model ([23-28]).
- **Model 2a:** Two-factor structure with three somatic items-sleep disturbance, fatigue, appetite changes [15,29-31].
- **Model 2b:** Two-factor structure adding psychomotor difficulties to the somatic dimension [15,29,30].
- **Model 2c:** Two-factor structure adding difficulties in concentration as a fifth somatic item (Table 2).

Model fit was evaluated using Root Mean Squared Error of Approximation (RMSEA), Tucker-Lewis Index (TLI) and Comparative Fit Index (CFI). Fit criteria were as follows: RMSEA <0.08 (exact fit = 0.00, close fit = 0.01-0.05, acceptable fit = 0.05-0.08, mediocre fit = 0.08-0.10, and poor fit = greater than 0.10), and TLI and CFI >0.90 (exact fit = 1.00, close fit = 0.95-0.99, acceptable fit = 0.90-0.95, mediocre fit = 0.85-0.90, and poor fit = less than 0.85) [39]. To improve fit, modification indices were consulted where necessary [40].

To determine measurement invariance, we followed the procedures outlined by Gregorich (2006) and Sass

(2014) [41,42]. Consistent with the recommendations of Cheung and Rensvold (2002); [7] and Chen (2007); [6], measurement invariance at each step was evaluated based on a change < -0.01 in CFI and < 0.015 in RMSEA. When all four tests are satisfied, invariance was established, and the most parsimonious model was retained.

Construct validity was assessed by evaluating convergent and divergent validity using the depression and anxiety subscales of the locally validated Depression, Anxiety, and Stress Scale (DASS) (Rekha, 2012) [43].

## Results

### PHQ-9 factor structure

Model 1, the single factor model, did not demonstrate acceptable fit, even after adjustments based on modification indices (See Table 2). Among the two-factor models, Model 2a, which specified three somatic items, provided the best model fit and was retained as the baseline model for subsequent analyses. See Figure 1 for factor loadings. Reliabilities of the PHQ-9 cognitive/affective and somatic dimensions were acceptable, with Cronbach's alpha values of 0.79 and 0.82, respectively.

**Table 2. Confirmatory factor analyses of the PHQ-9 – Models and fit indices**

	<i>Model 1</i>	<i>Model 2a</i>	<i>Model 2b</i>	<i>Model 2c</i>
1. Anhedonia	Depression	Cognitive/Affective	Non somatic	Non somatic
2. Depressed mood	Depression	Cognitive/Affective	Non somatic	Non somatic
3. Sleep disturbance	Depression	Somatic	Somatic	Somatic
4. Fatigue	Depression	Somatic	Somatic	Somatic
5. Appetite changes	Depression	Somatic	Somatic	Somatic
6. Low self-esteem	Depression	Cognitive/Affective	Non somatic	Non somatic
7. Concentration difficulties	Depression	Cognitive/Affective	Non somatic	Somatic
8. Psychomotor disturbances	Depression	Cognitive/Affective	Somatic	Somatic
9. Suicidal ideation	Depression	Cognitive/Affective	Non somatic	Non somatic
$\chi^2$	924.27	420.87	499.76	510.78
df	59	50	50	50
p-Value	<.001	<.001	<.001	<.001
CFI	0.835	0.985	0.897	0.892
TLI	0.789	0.978	0.875	0.871
RMSEA	0.067	0.041	0.052	0.052

Note:  $\chi^2$  = Chi-square value, df = Degrees of freedom; CFI = Comparative fit index; TL I= Tucker-Lewis index; RMSEA = Root mean square error of approximation.

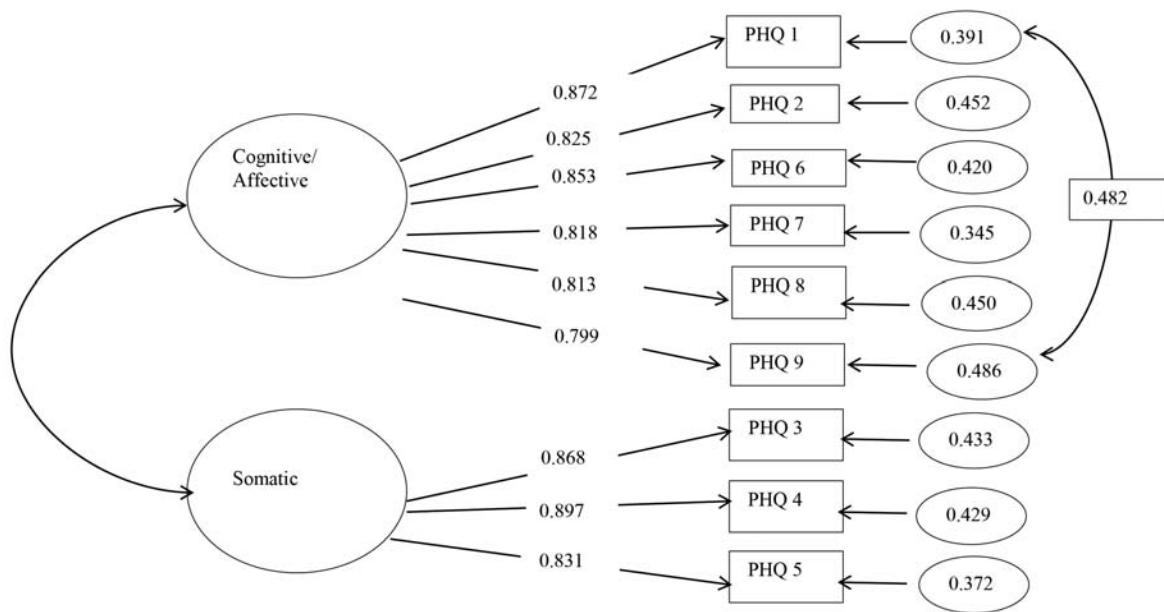


Figure 1. Two Factor model of the PHQ-9.

### Invariance across the two language instruments

The baseline model, which simultaneously tested the PHQ-9 using data from the Sinhala and Tamil versions, showed close fit (Table 3), supporting configural invariance across the two translations. Scalar and strict invariance were also established, as fit statistics showed minimal change when factor loadings and item thresholds were equated and when the error terms were equated across the two languages ( $\Delta\text{CFI} < 0.01$ ,  $\Delta\text{RMSEA} < 0.015$ ; Table 4).

These findings confirm that the Sinhala and Tamil versions of the PHQ-9 as used by the SLHAS can be treated as equivalent. For subsequent analyses, data from both versions were pooled.

### Invariance across sub-groups: gender, ethnicity, education level and socioeconomic status

The procedure to establish configural, scalar and strict invariance across gender, ethnicity, education level and socioeconomic status demonstrated close fit at all steps of invariance testing (Table 4). All models satisfied the  $\Delta\text{CFI} (< -0.01)$  and  $\Delta\text{RMSEA} (< 0.015)$  criteria, signifying invariance across all groups. Further, these findings held even with the PHQ-8 scale, which omits the last PHQ-9 item (i.e., suicidal ideation). This suggests that the PHQ-8 is a reliable and equivalent alternative for

research scenarios where a more concise measure is preferred.

### Robustness analysis of wording of item 8

In SLHAS Wave 1, items 6, 7 and 8 of the PHQ-9 were truncated. For item 8, for instance, the final words of the original phrasing were omitted ("Moving or speaking so slowly that other people could have noticed"). Despite this, Figure 1 shows that the item's loading was comparable to the other items, and the model retained a well-defined structure [44]. To further examine if the truncated wording affected item performance, we randomized administration of the truncated and the corrected, complete versions of the item during the SLHAS Wave 2x (June-September 2023). A total of 1,964 respondents were surveyed, with 1,027 receiving the truncated version and 937 the corrected, complete version. Response patterns were analyzed using the previously described methods, and no significant differences were observed ( $\Delta\text{CFI} < -0.01$ ;  $\Delta\text{RMSEA} < 0.015$ ). Additionally, a logistic regression, controlling for participant characteristics (gender, age, language, ethnicity, socioeconomic status, and education level) found no significant effect of wording on the likelihood of a positive response to this item (OR 0.98; 95% CI: 0.62-1.56). These findings indicate that the truncation did not materially impact responses, allowing us to conclude that the error had no significant effect on the results.

**Table 3. Testing dimensional invariance of the PHQ-9 Model 2a – fit indices across translated versions (Sinhala and Tamil), gender, ethnicity, levels of education and socio-economic status**

	$\chi^2$	df	p-Value	CFI	TLI	RMSEA
<b>Language of the questionnaire (n=6,346)</b>						
Sinhala	128.48	26	<.001	0.987	0.959	0.050
Tamil	233.21	28	<.001	0.962	0.931	0.054
<b>Gender (n=6,346)</b>						
Male	200.89	26	<.001	0.981	0.950	0.050
Female	219.77	26	<.001	0.971	0.945	0.050
<b>Ethnicity (n=6,346)</b>						
Sinhala	250.21	25	<.001	0.989	0.963	0.047
Sri Lankan Tamil	133.76	26	<.001	0.970	0.947	0.052
Muslim	42.44	26	<.001	0.983	0.958	0.052
Indian Tamil	68.28	26	<.001	0.929	0.899	0.058
Other	52.89	26	<.001	0.898	0.872	0.063
<b>Education Level (n=6,346)</b>						
No schooling	78.78	26	<.001	0.911	0.892	0.063
Grade 1-5	80.98	26	<.001	0.978	0.942	0.050
Grade 6-12	112.21	26	<.001	0.998	0.977	0.042
Passed O/L	140.54	26	<.001	0.991	0.973	0.050
Passed A/L	115.72	26	<.001	0.959	0.922	0.054
Degree & above	19.87	26	<.001	1.000	0.984	0.031
<b>Socio-Economic Status (n=6,346)</b>						
Poorest	100.84	26	<.001	0.988	0.969	0.046
Poorer	99.82	26	<.001	0.985	0.956	0.048
Middle	127.21	25	<.001	0.975	0.952	0.050
Richer	85.27	26	<.001	0.974	0.948	0.054
Richest	84.88	26	<.001	0.992	0.972	0.048

Note:  $\chi^2$  = Chi-square value, df = Degrees of freedom, CFI = Comparative fit index, TLI = Tucker-Lewis index, RMSEA = Root mean square error of approximation.

### Convergent and divergent validity

The correlations between the PHQ-9 somatic and cognitive/affective dimensions and the DASS depression and anxiety scales demonstrated expected patterns. Both PHQ-9 dimensions showed stronger correlations with the DASS depression scale ( $r = .76$  and  $.78$ , respectively,

$p < .01$ ) compared to the anxiety scale ( $r = .19$  and  $.18$ , respectively,  $p < .05$ ) [45]. Further, the reliabilities (Cronbach's Alpha) of the PHQ-9 cognitive/affective and somatic dimensions, and the DASS-depression and anxiety dimensions were 0.79, 0.82, 0.88 and 0.86 respectively.

Table 4. Testing different invariances of the PHQ-9 Model 2a – fit indices across translated versions (Sinhala and Tamil), gender, ethnicity, and levels of educational and socio-economic status and indicators of change in fit

	$\chi^2$	df	$\Delta\chi^2$	p-Value	CFI	TLI	RMSEA	$\Delta CFI$	$\Delta RMSEA$
<b>Language of the questionnaire (n=6,346)</b>									
Configural	546.78	52	-	-	0.990	0.971	0.033	-	-
Metric	632.24	56	85.46	<.001	0.985	0.962	0.037	-0.005	0.004
Scalar	737.58	60	105.34	<.001	0.977	0.955	0.040	-0.008	0.003
Strict	883.19	76	145.61	<.001	0.970	0.949	0.042	-0.007	0.002
<b>Gender (n=6,346)</b>									
Configural	490.84	52	-	-	0.987	0.960	0.036	-	-
Metric	528.92	62	38.08	<.001	0.985	0.962	0.037	-0.002	0.001
Scalar	623.06	71	94.14	<.001	0.985	0.963	0.038	0.000	-0.001
Strict	725.18	83	102.12	<.001	0.982	0.964	0.038	-0.003	-0.000
<b>Ethnicity (n=6,346)</b>									
Configural	528.47	98	-	-	0.982	0.966	0.030	-	-
Metric	600.88	128	72.41	<.001	0.978	0.960	0.027	-0.004	-
Scalar	725.95	147	125.07	<.001	0.972	0.959	0.023	-0.006	-0.004
Strict	853.16	188	127.21	<.001	0.972	0.958	0.023	-0.000	0.000
<b>Education Level (n=6,346)</b>									
Configural	673.73	132	-	-	0.978	0.937	0.022	-	-
Metric	780.55	190	106.82	<.001	0.970	0.936	0.021	-0.008	-0.001
Scalar	911.28	252	130.73	<.001	0.964	0.932	0.020	-0.006	-0.001
Strict	1040.90	295	129.62	<.001	0.960	0.948	0.021	-0.004	0.001
<b>Socio-Economic Status (n=6,346)</b>									
Configural	540.55	121	-	-	0.983	0.945	0.023	-	-
Metric	588.83	154	48.28	<.001	0.977	0.951	0.022	-0.006	-0.001
Scalar	677.52	189	88.69	<.001	0.973	0.949	0.019	-0.004	-0.003
Strict	798.11	237	120.59	<.001	0.965	0.943	0.020	-0.008	0.001

Note.  $\chi^2$  = Chi-square value; df = Degrees of freedom; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation;  $\Delta\chi^2$  = Change of Chi-square value;  $\Delta CFI$  = Change of comparative fit index;  $\Delta RMSEA$  = Change of Root mean square error of approximation.

## Discussion

Our findings indicate that the two-factor model, comprising somatic and cognitive/affective dimensions [see also 15, 29-31] and with three somatic items, works best for scoring PHQ-9 data in a representative Sri Lankan adult population, rather than the simpler one-factor solution. While studies using the PHQ-9 in Sri Lanka [10,12,19-21] have tended to report a single score for depression, our results suggest that reporting separate somatic and cognitive/affective scores may add value and support more nuanced understanding of depressive symptoms [15].

Measurement invariance testing indicates that the PHQ-9 and its cognitive/affective and somatic components perform equivalently in Sri Lanka, across Sinhala and Tamil language versions, as well as across gender, ethnicity, education level, and socioeconomic status. This supports the comparability of observed PHQ-9 means and variances/covariances across these subgroups, minimizing the risk of bias [41]. In other words, the PHQ-9 enables meaningful comparisons of depressive symptoms across languages and across major sociodemographic strata in Sri Lanka. Similar findings for the PHQ-8 suggest that it is also a robust and reliable alternative for certain scenarios where the suicidal ideation question cannot be asked.

The PHQ-9 also demonstrated strong validity, with robust support for its factor structure, measurement invariance, and convergent and divergent validity. These findings affirm the utility of the PHQ-9 and PHQ-8 as reliable tools for assessing depression in Sri Lanka generally, and specifically in the SLHAS. Future research using these validated tools can support reliable estimations of prevalence of depression in Sri Lanka and inform research on mental health interventions.

## Conclusion

This study provides strong evidence that the PHQ-9 (and the shorter PHQ-8 version) demonstrate measurement invariance across languages and sociodemographic groups in Sri Lanka. These findings enable reliable and unbiased comparisons of depressive symptoms across diverse subpopulations, laying the groundwork for future research on mental health prevalence. Given the country's recent challenges, including political and economic crises that exacerbated mental health issues [35], providing validated tools to assess and understand these problems is a critical first step. This study establishes a robust foundation for such future research.

## Author contributions

Sasanka worked on the analyses and initial drafts of the manuscript, Kumar was responsible for the overall supervision of the project and the finalising of the manuscript, while Rannan-Eliya provided feedback at all stages of the process.

## Conflict of interests

There are no conflict of interests.

## Acknowledgements

This study used data from the SLHAS. The SLHAS Wave 1 data collection was funded by the Swiss Agency for Development Cooperation (SDC) and the Swiss National Science Foundation (SNSF) through the Swiss Programme for Research on Global Issues for Development (r4d programme) by the grant “Inclusive social protection for chronic health problems” (Grant number 400640\_160374), and the Institute for Health Policy Public Interest Research Fund (Grant number PIRF-2018-02), and the SLHAS Wave 2x data collection that supported this study was partly funded by the National Institute for Health and Care Research (NIHR) Global Health Research programme (award number 17/63/47).

## Ethical approval (for original articles)

Ethics approval was obtained from the Sri Lanka Medical Association Ethical Review Committee (ERC/18-022) for the data that was used for this study. The research and all methods were accepted as conforming to relevant guidelines and regulations stipulated by the Committee. Study information was provided to all participants, together with an official letter from Ministry of Health, Sri Lanka, encouraging participation. Participants were, however, informed of their rights and consent was obtained prior to their participation (see Rannan-Eliya et al, 2022).

Rannan-Eliya RP, Wijemunige N, Perera P, Kapuge Y, Gunawardana N, Sigera C, *et al.* (2022). Prevalence and Associations of Hypertension in Sri Lankan Adults: Estimates from the SLHAS 2018-19 Survey Using JNC7 and ACC/AHA 2017 Guidelines. *Glob Heart*. 2022; 17(1).

## Patient consent (for case reports where the material is identifiable)

Not applicable.

## Sources of funding (if relevant)

Swiss Agency for Development Cooperation (SDC) and the Swiss National Science Foundation (SNSF) through the Swiss Programme for Research on Global Issues for Development (r4d programme) by the grant “Inclusive Social Protection for Chronic Health Problems” (Grant number 400640\_160374), and the Institute for Health Policy Public Interest Research Fund (Grant number PIRF-2018-02).

## Abbreviations (where relevant)

PHQ-9 – Patient Health Questionnaire-9  
 SLHAS – Sri Lanka Health and Ageing Study  
 NMHS – National Mental Health Survey  
 CFA – Confirmatory Factor Analysis  
 W1 – Wave 1  
 W2x – Extended Wave 2  
 CAPI – Computer-Assisted Personal Interview (CAPI)  
 DASS – Depression, Anxiety, and Stress Scale  
 O/L – Ordinary Level Examinations  
 A/L – Advanced Level Examinations  
 KMO – Kaiser-Meyer-Olkin  
 ULS – Unweighted Least Squares Estimation  
 SPSS AMOS – Statistical Package for Social Sciences – Analysis of Moment Structures  
 RMSEA – Root Mean Squared Error of Approximation  
 TLI – Tucker-Lewis Index  
 CFI – Comparative Fit Index

## References

- Alwis I, Baminiwatta A, Chandradasa M. Prevalence and associated factors of depression in Sri Lanka: A systematic review and meta-analysis. *Social Psychiatry and Psychiatric Epidemiology* 2023; **59**(2): 353-73. doi:10.1007/s00127-023-02495-z
- Institute For Research and Development. NMHS Final report - 2007. 2007; (May): 1-74.
- Sousa VD, Rojjanasirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *J Eval Clin Pract.* 2011; **17**(2): 268-74.
- Levis B, Yan XW, He C, Sun Y, Benedetti A, Thombs BD. Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: A meta-research review. *BMC Med.* 2019; **17**(1): 1-10.
- Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Dev Rev [Internet].* 2016; **41**: 71-90. Available from: <http://dx.doi.org/10.1016/j.dr.2016.06.004>
- Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model.* 2007; **14**(3): 464-504.
- Cheung GW, Rensvold RB. Structural Equation Modeling: A Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Struct Equ Model A Multidiscip J.* 2009; **9**(2): 233-55.
- Ball HA, Siribaddana SH, Kovas Y, Glozier N, McGuffin P, Sumathipala A, *et al.* Epidemiology and symptomatology of depression in Sri Lanka: A cross-sectional population-based survey in Colombo District. *J Affect Disord.* 2010; **123**(1-3): 188-96.
- Rajapakse OBW, Sivayogan S, Kulatunga PM. Prevalence and correlates of depression among older urban community-dwelling adults in Sri Lanka. *Psychogeriatrics.* 2019; **19**(3): 202-11.
- Senarath U, Wickramage K, Peiris SL. Prevalence of depression and its associated factors among patients attending primary care settings in the post-conflict Northern Province in Sri Lanka: A cross-sectional study. *BMC Psychiatry.* 2014; **14**(1): 1-10.
- Malhotra R, Chan A, Østbye T. Prevalence and correlates of clinically significant depressive symptoms among elderly people in Sri Lanka: Findings from a national survey. *Int Psychogeriatrics.* 2010; **22**(2): 227-36.
- Amarasuriya SD, Jorm AF, Reavley NJ. Prevalence of depression and its correlates among undergraduates in Sri Lanka. *Asian J Psychiatry [Internet].* 2015; **15**: 32-7. Available from: <http://dx.doi.org/10.1016/j.ajp.2015.04.012>
- Wickramasinghe A, Essén B, Surendhirakumaran R, Axemo P. Prevalence of depression among students at a Sri Lankan University: A study using the Patient Health Questionnaire-9 (PHQ-9) during the COVID-19 pandemic. *BMC Public Health.* 2023; **23**(1): 1-9.
- Kroenke K, Spitzer RL, Williams JBW. The PHQ-9.pdf. *J Gen Intern Med.* 2001; **16**: 606-13.
- Patel JS, Youngha, Rand KL, Wu. W, Cyders MA, Kroenke K. SJC. Measurement Invariance of the Patient Health Questionnaire-9 (PHQ-9) Depression Screener in U.S. Adults Across Sex, Race/Ethnicity and Educational Level: NHANES 2005-2016. *Physiol Behav.* 2018; **176**(1): 139-48.
- Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, *et al.* Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review. *J Affect Disord.* 2021; **279**(November): 473-83.
- Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatr Ann.* 2002; **32**(9): 509-15.
- Kalpajian CZ, Toussaint LL, Albright KJ, Bombardier CH, Krause JK, Tate DG. Patient health questionnaire-9 in spinal cord injury: An examination of factor structure as related to gender. *J Spinal Cord Med.* 2009; **32**(2): 147-56.
- Kulathilaka S, Hanwella R, de Silva VA. Depressive disorder and grief following spontaneous abortion. *BMC Psychiatry [Internet].* 2016; **16**(1): 1-6. Available from: <http://dx.doi.org/10.1186/s12888-016-0812-y>
- Wickramasinghe ND, Wijesinghe PR, Dharmaratne SD, Agampodi SB. The prevalence and associated factors of depression in policing: a cross sectional study in Sri Lanka. *Springerplus.* 2016; **5**(1).

21. Amarasuriya SD, Jorm AF, Reavley NJ. Depression literacy of undergraduates in a non-western developing context: the case of Sri Lanka Psychiatry. *BMC Res Notes*. 2015; **8**(1): 1-14.
22. Hanwella R, Ekanayake S, De Silva VA. The validity and reliability of the Sinhala translation of the patient health questionnaire (PHQ-9) and PHQ-2 screener. *Depress Res Treat*. 2014; 2014.
23. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Br J Gen Pract*. 2008; **58**(546): 32-6.
24. González-Blanch C, Medrano LA, Muñoz-Navarro R, Ruíz-Rodríguez P, Moriana JA, Limonero JT, et al. Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS One*. 2018; **13**(2): 1-16.
25. Cameron IM, Crawford JR, Lawton K, Reid IC. Differential item functioning of the HADS and PHQ-9: An investigation of age, gender and educational background in a clinical UK primary care sample. *J Affect Disord* [Internet]. 2013; **147**(1-3): 262-8. Available from: <http://dx.doi.org/10.1016/j.jad.2012.11.015>
26. Dum M, Pickren J, Sobell LC, Sobell MB. Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. *Addict Behav*. 2008; **33**(2): 381-7.
27. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med*. 2006; **21**(6): 547-52.
28. Kocalevent RD, Hinz A, Brähler E. Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. *Gen Hosp Psychiatry* [Internet]. 2013; **35**(5): 551-5. Available from: <http://dx.doi.org/10.1016/j.genhosppsych.2013.04.006>
29. Petersen JJ, Paulitsch MA, Hartig J, Mergenthal K, Gerlach FM, Gensichen J. Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. *J Affect Disord* [Internet]. 2015; **170**: 138-42. Available from: <http://dx.doi.org/10.1016/j.jad.2014.08.053>
30. Krause JS, Bombardier C, Carter RE. Assessment of Depressive Symptoms During Inpatient Rehabilitation for Spinal Cord Injury: Is There an Underlying Somatic Factor When Using the PHQ? *Rehabil Psychol*. 2008; **53**(4): 513-20.
31. Elhai JD, Contractor AA, Tamburrino M, Fine TH, Prescott MR, Shirley E, et al. The factor structure of major depression symptoms: A test of four competing models using the Patient Health Questionnaire-9. *Psychiatry Res*. [Internet]. 2012; **199**(3): 169-73. Available from: <http://dx.doi.org/10.1016/j.psychres.2012.05.018>
32. Rannan-Eliya RP, Dissanayake VH, Perera P, Perera B, Herath HMM, Wijemunige N, Fernando EY. Cohort Profile: The Sri Lanka Health and Ageing Study (SLHAS). *International Journal of Epidemiology* 2024; **53**(2): doi:10.1093/ije/dyae044
33. Rannan-Eliya RP, Wijemunige N, Perera P, Kapuge Y, Gunawardana N, Sigera C, et al. Prevalence and Associations of Hypertension in Sri Lankan Adults: Estimates from the SLHAS 2018-19 Survey Using JNC7 and ACC/AHA 2017 Guidelines. *Glob Heart*. 2022; **17**(1).
34. Rannan-Eliya RP, Wijemunige N, Perera P, Kapuge Y, Gunawardana N, Sigera C, et al. Prevalence of diabetes and pre-diabetes in Sri Lanka: A new global hotspot-estimates from the Sri Lanka Health and Ageing Survey 2018/2019. *BMJ Open Diabetes Res Care* 2023; **11**(1): 1-11.
35. Deivanayagam TA, Chobhthaigh S, Devakumar D, et al. Mental health prevalence, healthcare use and access between 2018 and 2022 in Sri Lanka: an analysis of survey data. *Global Health Research* 2024. doi: 10.3310/HJWA5078
36. Kiwanuka F, Kopra J, Sak-Dankosky N, Nanyonga RC, Kvist T. Polychoric Correlation with Ordinal Data in Nursing Research. *Nurs Res*. 2022; **71**(6): 469-76.
37. Lorenzo-Seva U, Ferrando PJ. Lorenzo-Seva-Ferrando. Article FACTORA Computer Program To Fit The-2. *Behav Res Methods*. 2006; **38**(1): 88-91.
38. Katsikatsou M, Moustaki I, Yang-Wallentin F, Jöreskog KG. Pairwise likelihood estimation for factor analysis models with ordinal data. *Comput Stat Data Anal* [Internet]. 2012; **56**(12): 4243-58. Available from: <http://dx.doi.org/10.1016/j.csda.2012.04.010>
39. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model*. 1999; **6**(1): 1-55.
40. Byrne BM. Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *J Pers Assess*. 2005; **85**(1): 17-32.
41. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care*. 2006; **44**(11 SUPPL. 3): 15-9.
42. Sass DA, Schmitt TA, Marsh HW. Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Struct Equ Model*. 2014; **21**(2): 167-80.
43. Rekha AVS. Adaptation and validation of the Depression, Anxiety, and Stress Scale (DASS21) among Students of the University of Colombo. Annual Research Symposium, University of Colombo. 2012.
44. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 2010.
45. Ferketich SL, Figueredo AJ, Knapp TR. The multitrait-multimethod approach to construct validity. *Research in nursing & health*, 1991; **14**(4): 315-20. Available from <https://doi.org/10.1002/nur.4770140410>