

Research paper

Examining the psychometric properties of the Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7 among young urban South African women

Claire Hart^{a,*}, Catherine E. Draper^a, Larske M. Soepnel^{a,b}, Motlatso Godongwana^a, Khuthala Mabetha^a, Lukhanyo H. Nyati^{a,c}, Simone H. Crouch^a, Shane A. Norris^{a,d}

^a MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, School of Clinical Medicine, University of the Witwatersrand, Johannesburg, South Africa

^b Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

^c Interprofessional Education Unit, Faculty of Community and Health Sciences, University of the Western Cape, Cape Town, South Africa

^d School of Human Development and Health, University of Southampton, Southampton, UK

ARTICLE INFO

Keywords:

Patient Health Questionnaire
Depression
Generalized anxiety disorder
Anxiety
Psychometric properties
South Africa

ABSTRACT

Background: Valid, reliable, and easy-to-administer scales are crucial for identifying mental health conditions, especially in LMICs where such scales tend not to be validated. This study aims to address this gap by investigating the psychometric properties and factorial structure of the PHQ-9 and GAD-7 in a sample of young women in Soweto, South Africa.

Methods: The PHQ-9 and GAD-7 were administered to 6028 women aged 18–28 years old. Cronbach's alpha, Mokken scale analysis, and Confirmatory Factor Analysis were used to provide support for the internal consistency and construct validity of these scales.

Results: Both scales demonstrated good internal consistency ($\alpha = 0.81$ for PHQ-9 and $\alpha = 0.84$ for GAD-7). Internal consistency reliability was further supported by positive inter-item correlations and item-by-scale correlations for all items on both measures. CFA of the PHQ-9 and GAD-7 showed a reasonable fit for the 1-factor model and 2-factor models (depression and anxiety with somatic and cognitive subtypes).

Limitations: This study was limited to young African women in urban Soweto who were proficient in English, which may affect generalizability. Differences in language or cultural context may impact the accuracy and applicability of these scales to other African populations.

Conclusion: The PHQ-9 and GAD-7 are valid and reliable for identifying psychological distress in the studied population. Despite showing good psychometric properties, further diagnostic assessment is needed to confirm clinical diagnoses. The scales are useful for identifying those at risk but not a substitute for comprehensive diagnostic evaluations.

1. Introduction

Depression and anxiety are universally prevalent chronic non-communicable diseases (NCD) with a significant contribution to the Global Burden of Diseases (Santomauro et al., 2021). Depression is classified as a prevailing sense of sadness or hopelessness, whilst anxiety involves an overwhelming sense of worry, nervousness, and fear (APA, 2013). The widespread prevalence of depression (26 %) and anxiety (18 %) in South African adults have been estimated in a large-scale nationally representative weighted survey conducted in 2021 (Craig et al.,

2022). Women have been shown to have a higher risk of probable depression compared to men (Craig et al., 2022; Jewkes et al., 2010), with nearly one in every five women reporting moderate to severe levels of anxiety (Mkhwanazi and Gibbs, 2021). Among young women in Africa, depression and anxiety are the leading causes of disability, morbidity, and mortality (Materu et al., 2020). With approximately 19 % and 15 % of young South African women reporting symptoms of probable depression and anxiety (Draper et al., 2022a, 2022b; Draper et al., 2023; Redinger et al., 2020), mental health should be prioritized as a public health concern for this population.

* Corresponding author at: PO BOX 21444, Helderkruijn 1733, South Africa.
E-mail address: claire.hart@wits.ac.za (C. Hart).

<https://doi.org/10.1016/j.jad.2024.09.145>

Received 7 March 2024; Received in revised form 16 September 2024; Accepted 21 September 2024

Available online 23 September 2024

0165-0327/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Depression and anxiety are distinctive mental health conditions; however, a combination of depressive and anxiety symptoms can warrant the diagnosis of probable depression and anxiety at the same time (APA, 2013). This is further complicated by the overlap of certain symptoms in both conditions, such as increased irritability, changes in sleep and energy levels, and cognitive difficulties, which can make it difficult to distinguish between the two conditions. The presence of comorbid depression and anxiety complicates treatment outcomes (Vittengl et al., 2019), highlighting the importance of identifying both conditions before deciding on the necessary treatment. The comorbidity of depression and anxiety has been found to be significant in the South African context. One nationwide study showed that 58.1 % of South African adults with probable depression also reported probable anxiety (Craig et al., 2022). There is a higher severity risk and burden with comorbid depression and anxiety, including functional impairments, social and occupational difficulties, suicidal ideation and attempts, and a higher risk of developing other mental health conditions (Materu et al., 2020; Weiss et al., 2016). Due to the probable prevalence of comorbid depression and anxiety in the South African context, it is important to have appropriate scales to estimate different types of clinical profiles through the identification of patterns of mental health symptoms.

The routine administration of mental health scales in primary health care settings is recommended to ensure early detection, appropriate referral for treatment, and reduction of the burden of depression and anxiety (Kroenke et al., 2007; Materu et al., 2020; Stein et al., 2014). The Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) are self-administered scales based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for depression and anxiety (APA, 2013). The PHQ-9 and GAD-7 have been established to be reliable and valid scales for measuring depression and anxiety separately and as comorbid conditions (Kroenke et al., 2001; Spitzer et al., 2006). Whilst these scales were developed and evaluated in Europe and North America, they have been frequently used in settings with poor mental health service provision and a lack of specialized staff (Mughal et al., 2020). They have reliably been used for detecting depression and anxiety in South African adolescents (Marlow et al., 2022), as well as in other African countries (Materu et al., 2020; Monahan et al., 2009; Osborn et al., 2020). Nevertheless, there is a dearth of literature on the use of these mental health scales in low- and middle-income countries (LMICs) such as South Africa, and particularly among young women, which is the gap this study intends to address. This study aimed to examine the internal consistency reliability (how well the items on each scale measure the same construct and are consistent with each other?) and construct validity (does each scale accurately measure the theoretical construct it is supposed to assess?) of the PHQ-9 and GAD-7.

This study will also assess the factorial structure of the PHQ-9 and GAD-7 to determine the possibility of somatic and cognitive subdimensions for this population, as well as explore how these subdimensions could be related to presentations of depression and anxiety. Whilst the PHQ-9 and GAD-7 were developed as unidimensional scales for depression and anxiety (1-factor models), numerous confirmatory factor analysis (CFA) investigations have recently identified specific items reflecting a cognitive factor and somatic factor, providing evidence for a 2-factor model for each scale (Boothroyd et al., 2018; De Man et al., 2021; Lamela et al., 2020; Rutter and Brown, 2017). The PHQ-9 items which load on the cognitive factor are P1 (Anhedonia), P2 (Depressed mood), P6 (Feelings of worthlessness) and P9 (Suicidal ideation), whereas P3 (Sleep difficulties), P4 (Fatigue), P5 (Appetite changes), P7 (Concentration difficulties) and P8 (Psychomotor agitation) load on to the somatic factor. For the GAD-7, items which load on the cognitive factor are G1 (Feeling nervous, anxious or on edge), G2 (Cannot control worry), G3 (Worry about a variety of topics, events, or activities) whereas G7 (Feeling afraid as if something awful might happen), and G4 (Trouble relaxing), G5 (Restlessness) and G6 (Easily annoyed or irritable) load on to the somatic factor. Since the somatic and

cognitive subdimensions have been shown to correspond with different clinical presentations of depression and anxiety (Boothroyd et al., 2018; De Man et al., 2021; Lamela et al., 2020; Rutter and Brown, 2017), distinguishing between potential depression and anxiety subtypes can inform appropriate mental health interventions. Furthermore, specific item responses can be more prevalent for certain subgroups as item-specific variances have been found in other non-Western populations in Africa (Materu et al., 2020; Monahan et al., 2009; Osborn et al., 2020) and India (De Man et al., 2021). Examining the factorial structure of data from a specific population can elucidate the probable existence of depression and anxiety subdimensions that may be relevant to different cultures and settings (De Man et al., 2021).

The study aimed to provide evidence on whether the PHQ-9 and GAD-7 are psychometrically reliable and valid scales for use among 18- to 28-year-old women living in Soweto. Concerns regarding the application of these scales in different cultural and social contexts (De Man et al., 2021; Materu et al., 2020; Monahan et al., 2009; Osborn et al., 2020) drive this research. The study explored how these scales can accurately interpret mental health constructs in a sample of young South African women, providing evidence of their utility and generalizability across different populations. The diagnostic value of these instruments in recognizing depressive and anxiety disorders can only be determined if the scales are psychometrically applicable and valid for specifically defined age groups, cultures, and settings. Therefore, this research investigated the psychometric properties of the PHQ-9 and GAD-7 in this population to ensure the appropriateness of these mental health assessment scales, to improve evidence-based mental health care for women in South Africa.

2. Materials and methods

2.1. Population and setting: HeLTI SA Bukhali

This study formed part of the *Bukhali* randomised controlled trial as part of the Healthy Life Trajectories Initiative (HeLTI) aims to evaluate the impact of a complex continuum of care interventions for young women aged between 18 and 28 years old (Norris et al., 2022). The *Bukhali* intervention focuses on enhancing physical and mental health to establish healthier trajectories for women and their children (Draper et al., 2023). This study is being conducted in Soweto, which is a densely populated urban setting with a predominantly low- to middle-income and multilingual population in Johannesburg. Participants were recruited through a survey of approximately 30,000 Soweto households through community-based recruitment methods described in detail elsewhere (Norris et al., 2022). Young women in Soweto who were not pregnant and had no previous diagnosis of cancer, Type 1 Diabetes, or Epilepsy were invited to participate in the trial. If they were willing to participate, baseline data were collected from 6062 women during a single visit, including depression and anxiety indicators, which formed the focus of this study. Participants who were missing data on depression and anxiety indicators were excluded, thus, the included sample for this analysis was 6028 women.

2.2. Ethical considerations

Ethical approval for the HeLTI SA trial and process evaluation was obtained from the Human Research Ethics Committee (Medical) (HREC) of the University of the Witwatersrand, South Africa (reference: M1811111). The trial is registered with the Pan African Clinical Trials Registry (<https://pactr.samrc.ac.za> identifier: PACTR201903750173871).

2.3. Mental health scales

The PHQ-9 is routinely used to assess the presence and severity of depression symptoms (Kroenke et al., 2001). Participants were asked to

rate how often over the last two weeks they were bothered by 9 specific symptoms, with responses recorded on a four-point Likert scale ranging from 0 (“Not at all”) to 3 (“Nearly every day”). Based on these responses, the level of probable depression was categorized into five groups: minimal (scoring in the ranges of 0–4), mild (scoring in the ranges of 5–9), moderate (scoring in the ranges of 10–14), moderately severe (scoring in the ranges of 15–19), and severe (scoring in the ranges of 20–27). The unidimensional representation of probable depression was defined by a PHQ-9 score of 10 or greater.

The GAD-7 is routinely used to assess the presence and severity of anxiety symptoms (Spitzer et al., 2006). Participants were asked to rate how often over the last two weeks they were bothered by 7 specific symptoms, with responses recorded on a four-point Likert scale ranging from 0 (“Not at all”) to 3 (“Nearly every day”). Based on these self-report responses, the level of probable anxiety was categorized into four groups: minimal (scoring in the ranges of 0–4), mild (scoring in the ranges of 5–9), moderate (scoring in the ranges of 10–14), and severe (scoring in the ranges of 15–21). The unidimensional representation of probable anxiety was defined by a GAD-7 score of 10 or greater.

Since cut-off values for the general South African population or specific population groups have not been established, we used the categories offered by the PHQ-9 authors (Kroenke et al., 2001) and the GAD-7 authors (Spitzer et al., 2006). The cut-off values were used to represent the prevalence of probable depression (PHQ-9 score of 10 or greater), probable anxiety (GAD-7 score of 10 or greater) and probable comorbid depression and anxiety (PHQ-9 score of 10 or greater AND GAD-7 score of 10 or greater).

2.4. Statistical analysis

Data were cleaned and analyzed using Stata® (Version 17.0, Stata-Corp, College Station, TX, USA). The PHQ-9 and the GAD-7 were screened for missing values and normality, assessed at the univariate and multivariate levels. The distribution of the continuous variables was assessed using the Shapiro-Wilk test. All continuous variables were presented as means (standard deviation, SD) and categorical variables were presented as frequencies (%).

To determine the scale reliability of both the PHQ-9 and GAD-7, Cronbach's alpha was used, where α value of 0.70 or greater was considered acceptable, given that a higher α value tends to indicate higher internal consistency (Streiner, 2003). Since the PHQ-9 and GAD-7 were developed as unidimensional measures and the length of a measure has been shown to affect α , item characteristics and discriminations were also explored to provide additional evidence for internal consistency reliability (Streiner, 2003). Item-means, inter-item correlations, and item-by-scale correlations were investigated, with acceptable inter-item correlations considered to fall within the range of 0.20 to 0.50 and item-by-scale correlations to fall within the range of 0.20 to 0.39 (Clark and Watson, 1995).

To further understand the dimensional structure of the measures, a non-parametric method of item response theory approach called Mokken analysis was adopted to investigate the number of items measuring the same underlying concept in each measure (Mokken, 1971). Loevinger's coefficient (H) was the main calculation in Mokken scale analysis, which revealed the extent to which pairs of items conformed to Guttman criteria. Although based on Guttman scaling, Mokken does not assume error-free data but is rather concerned with how scores on pairs of items were consistently relative to one another. Loevinger's H calculates the size of this error for each item, pairs of items, and the overall measure. Using the automated item selection procedure, the default scalability criterion was set at H of 0.3 or greater for each item. Loevinger's H (the “H value”) of 0.5 indicated a strong scale; weak scales were represented by H values of 0.3 and below.

A selection of models was tested using confirmatory factor analysis (CFA) to evaluate what model best suited the data. For the PHQ-9, we tested a 1-factor model (depression factor) and a correlated 2-factor

model (cognitive and somatic factors) as this has been proposed in prior studies (Lamela et al., 2020). For the GAD-7, we tested a 1-factor model (anxiety factor) and a correlated 2-factor model (cognitive and somatic factors) as this has been proposed in prior studies (Beard and Björgvinsson, 2014; Rutter and Brown, 2017; Doi et al., 2018). As the multivariate normality assumption was violated, the assessment of these models was based on their χ^2 -values, the item indicators' loadings and the following sample-corrected for non-normal data goodness of fit indices. Model fit was compared using the following indices and target cut-offs: the Standardized Root Mean Square Residual (SRMR) (≤ 0.08), the Root Mean Square Error of Approximation (RMSEA) with a 90 % confidence interval (≤ 0.08), Comparative Fit Index (CFI) (≥ 0.95), and the Tucker-Lewis Index (TLI) (≥ 0.95).

3. Results

3.1. Study samples description

Table 1 provides the demographic characteristics of the sample of 6028 women who participated in the study. The average of participants was 22.29 years (SD = 2.89), with an age range between 18 and 28 years of age. The majority of women had completed matric education (65.18 %), with a medium household asset score of 7.97 (SD = 2.06). Forty-six percent of women were single, 48.32 % were in committed relationships, 4.84 % were co-habiting with their partners, and 0.98 % were married. Age, household asset score and education level were fairly consistent across diagnostic groups, with slight variations in means and similar ranges. A notable pattern was observed in relationship status, with those who were single having a higher proportion in the probable depression and anxiety groups compared to being in a relationship, cohabiting, and being married.

3.2. Prevalence of depressive and anxiety symptoms

Using the cut-offs proposed by the PHQ-9 and the GAD-7 authors, the proportion of participants with probable depression was 14 % (identified by PHQ-9 score of 10 or greater); probable anxiety was 12 % (identified by GAD-7 score of 10 or greater); and probable comorbid anxiety and depression was 9 % (identified by both PHQ-9 score of 10 or greater, and GAD-7 score of 10 or greater). Thirty-one percent of participants were classified as at-risk for depression (identified by PHQ-9 score ranging between 5 and 9), 29 % were classified as at-risk for anxiety (identified by GAD-7 score ranging between 5 and 9), and 17 % were at-risk for probable comorbid depression and anxiety (identified by both PHQ-9 score ranging between 5 and 9, and GAD-7 score ranging between 5 and 9). Fifteen percent of participants were identified at risk for suicidality and/or self-harm, as indicated by agreement with the statement “Thoughts that you would be better off dead or of hurting yourself in some way”. These results are tabulated in Table 2.

3.3. Reliability

3.3.1. Unidimensional measures of depression and anxiety: PHQ-9 and GAD-7

The PHQ-9 demonstrated Cronbach's α of 0.81 and the GAD-7 demonstrated Cronbach's α of 0.84, suggesting both measures have robust reliability for identifying those at risk for depression and anxiety in this population (see Table 3). Inter-item correlations were positive (PHQ-9 average $r = 0.33$; GAD-7 average $r = 0.44$), as were item-by-scale correlations (PHQ-9 average $r = 0.63$; GAD-7 average $r = 0.72$). The two PHQ-9 items with the weakest scale correlations were items P8 (*Moving or speaking so slowly that other people could have noticed. Or the opposite*) and P7 (*Trouble concentrating on things, such as reading the newspaper or watching television*) with item-by-scale correlations of 0.57 and 0.59 respectively. These items also had the weakest individual item H values from the Mokken scale analysis of 0.32 and 0.34 respectively.

Table 1

Demographic characteristics of sample of Sowetan young women.

	Total		Probable depression		Probable anxiety		Probable co-morbid depression and anxiety	
	n	%/Mean (SD)	n	%/Mean (SD)	n	%/Mean (SD)	n	%/Mean (SD)
Age (years)	6028	22.29 (2.89), range 18–28	865	22.24 (2.86), range 18–28	741	22.24 (2.88), range 18–28	565	22.27 (2.88), range 18–28
Household asset score	6028	7.97 (2.06), range 0–13	865	7.92 (1.99), range 1–13	741	7.92 (2.07), range 1–13	565	7.91 (1.99), range 1–13
Level of education								
Not completed matric	2099	34.82	285	32.95	241	32.52	187	33.10
Completed matric	3929	65.18	580	67.05	500	67.48	378	66.90
Relationship status								
Single	2764	45.85	461	53.29	400	53.98	313	55.40
In a relationship	2913	48.32	359	41.50	304	41.03	221	39.12
Co-habiting	292	4.84	36	4.16	33	4.45	27	4.78
Married	59	0.98	9	1.04	4	0.54	4	0.71

Table 2

Prevalence of depression, anxiety, and suicidal ideation symptoms, and probable depression and anxiety diagnoses in Sowetan young women.

	Total	
	n =	%
	6028	
Depression symptoms		
Minimal (0–4)	3257	54.03
Mild (5–9)	1906	31.62
Moderate (10–14)	504	8.36
Moderately severe (15–19)	252	4.18
Severe (20–27)	109	1.81
Anxiety symptoms		
Minimal (0–4)	3545	58.81
Mild (5–9)	1742	28.90
Moderate (10–14)	448	7.43
Severe (15–21)	293	4.86
Suicidal ideation		
Thoughts that you would be better off dead or of hurting yourself in some way...		
Not at all	5115	84.85
Several days	540	8.96
More than half the days	124	2.06
Nearly every day	249	4.13
Probable diagnoses		
Probable depression	865	14.35
Probable anxiety	741	12.29
Probable depression and anxiety	565	9.37

The GAD-7 item with the weakest scale correlation was item G7 (*Feeling afraid as if something awful might happen*) with an item-by-scale correlation of 0.68. This item also had the weakest individual item H value of 0.43 from the Mokken scale analysis. All items on the PHQ-9 and GAD-7 demonstrated H values above the scalability criterion of 0.3 (PHQ-9 average H value = 0.37; GAD-7 average 0.49), suggesting that both measures are acceptable in terms of the internal consistency and scalability of items.

3.3.2. 2-Factor measures of depression and anxiety: PHQ-COG, PHQ-SOM and GAD-COG, GAD-SOM

To determine the possibility of cognitive and somatic subdimensions of depression in a South African population of young women, we separated the PHQ-9 into the PHQ-COG and the PHQ-SOM. The PHQ-COG demonstrated Cronbach's α of 0.74 and the PHQ-SOM demonstrated Cronbach's α of 0.67, suggesting the measure has adequate reliability for cognitive and somatic symptomology of depression (see Table 4). Inter-item correlations were positive (PHQ-COG average $r = 0.33$; PHQ-SOM average $r = 0.41$), as were item-by-scale correlations (PHQ-COG average $r = 0.66$; PHQ-SOM average $r = 0.78$). The two PHQ-COG items with the weakest scale correlations were items P8 (*Moving or speaking so slowly that other people could have noticed. Or the opposite*) and P7 (*Trouble concentrating on things, such as reading the newspaper or watching television*) with item-by-scale correlations of 0.62 for both items. These items also had the weakest individual item H values from the Mokken scale analysis of 0.35 and 0.33 respectively. The PHQ-SOM item with the weakest scale correlation was item P5 (*Poor appetite or overeating*) with an item-by-scale correlation of 0.76. This item also had the weakest individual item H value from the Mokken scale analysis of 0.40. All

Table 3

Reliability of unidimensional measures administered to young women in Soweto.

PHQ 1-factor model	Item-test corr.	Inter-item corr.	α	H	GAD 1-factor model	Item-test corr.	Inter-item corr.	α	H
P1 (<i>Little interest or pleasure in doing things</i>)	0.61	0.33	0.80	0.35	G1 (<i>Feeling nervous, anxious or on edge</i>)	0.70	0.44	0.83	0.47
P2 (<i>Feeling down, depressed, or hopeless</i>)	0.7	0.31	0.78	0.43	G2 (<i>Not being able to stop or control worrying</i>)	0.77	0.42	0.81	0.53
P3 (<i>Trouble falling or staying asleep or sleeping too much</i>)	0.65	0.33	0.79	0.37	G3 (<i>Worrying too much about different things</i>)	0.76	0.42	0.81	0.54
P4 (<i>Feeling tired or having little energy</i>)	0.67	0.32	0.79	0.39	G4 (<i>Trouble relaxing</i>)	0.73	0.43	0.81	0.50
P5 (<i>Poor appetite or overeating</i>)	0.63	0.33	0.80	0.36	G5 (<i>Being so restless that it is hard to sit still</i>)	0.70	0.44	0.83	0.49
P6 (<i>Feeling bad about yourself or that you are a failure or have let yourself down</i>)	0.64	0.33	0.79	0.38	G6 (<i>Becoming easily annoyed or irritable</i>)	0.70	0.45	0.83	0.47
P7 (<i>Trouble concentrating on things, such as reading the newspaper or watching television</i>)	0.59	0.34	0.80	0.33	G7 (<i>Feeling afraid as if something awful might happen</i>)	0.68	0.45	0.83	0.44
P8 (<i>Moving or speaking so slowly that other people could have noticed. Or the opposite</i>)	0.57	0.34	0.81	0.35					
P9 (<i>Thoughts that you would be better off dead or of hurting yourself in some way</i>)	0.63	0.33	0.80	0.40					
PHQ-9	0.63	0.33	0.81	0.37	GAD-7	0.72	0.44	0.84	0.49

Table 4

Reliability of 2-factor measures administered to young women in Soweto.

PHQ 2-factor model	Item-test corr.	Inter-item corr.	α	H	GAD 2-factor model	Item-test corr.	Inter-item corr.	α	H
P1 (<i>Little interest or pleasure in doing things</i>) Cognitive	0.64	0.33	0.72	0.36	G1 (<i>Feeling nervous, anxious or on edge</i>) Cognitive	0.64	0.33	0.74	0.48
P2 (<i>Feeling down, depressed, or hopeless</i>) Cognitive	0.73	0.30	0.68	0.44	G2 (<i>Not being able to stop or control worrying</i>) Cognitive	0.73	0.30	0.68	0.55
P3 (<i>Trouble falling or staying asleep or sleeping too much</i>) Somatic	0.78	0.41	0.58	0.42	G3 (<i>Worrying too much about different things</i>) Cognitive	0.78	0.41	0.69	0.57
P4 (<i>Feeling tired or having little energy</i>) Somatic	0.80	0.36	0.53	0.45	G4 (<i>Trouble relaxing</i>) Somatic	0.80	0.36	0.54	0.52
P5 (<i>Poor appetite or overeating</i>) Somatic	0.76	0.44	0.61	0.40	G5 (<i>Being so restless that it is hard to sit still</i>) Somatic	0.76	0.44	0.56	0.52
P6 (<i>Feeling bad about yourself or that you are a failure or have let yourself down</i>) Cognitive	0.69	0.32	0.70	0.41	G6 (<i>Becoming easily annoyed or irritable</i>) Somatic	0.69	0.32	0.72	0.46
P7 (<i>Trouble concentrating on things, such as reading the newspaper or watching television</i>) Cognitive	0.62	0.34	0.72	0.33	G7 (<i>Feeling afraid as if something awful might happen</i>) Cognitive	0.62	0.34	0.77	0.43
P8 (<i>Moving or speaking so slowly that other people could have noticed. Or the opposite</i>) Cognitive	0.62	0.34	0.72	0.35					
P9 (<i>Thoughts that you would be better off dead or of hurting yourself in some way</i>) Cognitive	0.67	0.32	0.70	0.41					
PHQ-COG	0.66	0.33	0.74	0.39	GAD-COG	0.66	0.33	0.78	0.51
PHQ-SOM	0.78	0.41	0.67	0.42	GAD-SOM	0.78	0.41	0.70	0.50

items demonstrated H values above the scalability criterion of 0.3 (PHQ-COG average H value = 0.39; PHQ-SOM average H value = 0.42), suggesting that this was a moderate measure in terms of the internal consistency and scalability of items.

To determine the possibility of cognitive and somatic subdimensions in the GAD-7, the GAD-7 was separated into the GAD-COG and GAD-SOM. The GAD-COG demonstrated Cronbach's α of 0.78 and the GAD-SOM demonstrated Cronbach's α of 0.70, suggesting the measure has good reliability for cognitive and somatic symptomology of anxiety (see Table 4). Inter-item correlations were positive (GAD-COG average $r = 0.47$; GAD-SOM average $r = 0.44$), as were item-by-scale correlations (GAD-COG average $r = 0.77$; GAD-SOM average $r = 0.79$). The GAD-COG item with the weakest scale correlation was item G7 (*Feeling afraid as if something awful might happen*) with an item-by-scale correlation of 0.72. This item also had the weakest individual item H value from the Mokken scale analysis of 0.43. The GAD-SOM item with the weakest scale correlation was item G6 (*Becoming easily annoyed or irritable*) with an item-by-scale correlation of 0.74. This item also had the weakest individual item H value from the Mokken scale analysis of 0.42. All items demonstrated H values above the scalability criterion of 0.3 (GAD-COG average H value = 0.39; GAD-SOM average H value = 0.41), suggesting that this was a moderate measure in terms of the internal consistency and scalability of items.

3.3.3. Validity of measures: confirmatory factor analysis

In addition to exploring the reliability of the measures, this study was interested in exploring whether the PHQ-9 and GAD-7 were valid measures to assess depression and anxiety in this population of young women. Univariate and multivariate normality was assessed for the items used in the various 1-factor and 2-factor models before we conducted the CFA. Since an estimation method robust to non-normality was required, the chi-square test and standard errors were adjusted using Satorra-Bentler robust standard errors. The fit indices, per factor model, are compared in Table 5.

CFA of the models proposed for the PHQ-9 demonstrated that the measures measured the constructs they were designed to measure, namely depression (1-factor model) and depression with cognitive and

Table 5

Fit statistics for proposed mental health models.

Model	X ²	df	p-Value	CFI	TLI	RMSEA	SRMR
Depression 1-factor PHQ-9	377.15	27	<0.05	0.95	0.93	0.05	0.03
Depression 2-factor PHQ-COG PHQ-SOM	238.81	26	<0.05	0.97	0.96	0.04	0.03
Anxiety 1-factor GAD-7	0.72	14	<0.05	0.94	0.90	0.08	0.04
Anxiety 2-factor GAD-COG GAD-SOM	510.75	13	<0.05	0.95	0.92	0.07	0.04

somatic symptomology (2-factor model). The correlation between the cognitive and somatic factors of the 2-factor model was high ($r = 0.86$). Factor loading estimates revealed that the indicators were strongly related to their purported factors for the 2-factor model (range $\lambda = 0.49$ –0.71). All items loaded significantly onto the single factor and loadings were also acceptable for the 1-factor model (range $\lambda = 0.48$ –0.69) (see Fig. 1). CFA revealed a better fit for a 2-factor model (CFI = 0.97, SRMR = 0.03, RMSEA = 0.04), whilst a 1-factor model was also acceptable (CFI = 0.95, SRMR = 0.03, RMSEA = 0.05) (see Table 5).

CFA of the models proposed for the GAD-7 demonstrated that the measures measured the constructs they were designed to measure, namely anxiety (1-factor model) and anxiety with cognitive and somatic symptomology (2-factor model). The correlation between the cognitive and somatic factors of the 2-factor model was high ($r = 0.88$). Factor loading estimates revealed that the indicators were strongly related to their purported factors (1-factor model range $\lambda = 0.59$ –0.76; 2-factor model range $\lambda = 0.58$ –0.78) (see Fig. 2). CFA revealed a better fit for a 2-factor model (CFI = 0.95, SRMR = 0.04, RMSEA = 0.07), whilst a 1-factor model was also acceptable (CFI = 0.94, SRMR = 0.04, RMSEA =

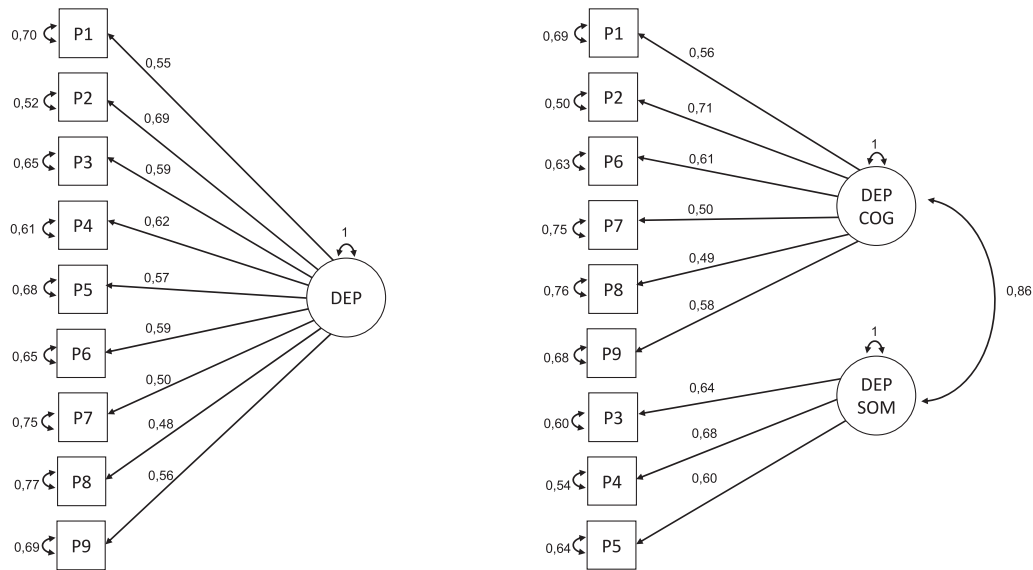


Fig. 1. Standardized factor loadings and error variances for 1- and 2-factor models of the PHQ-9.

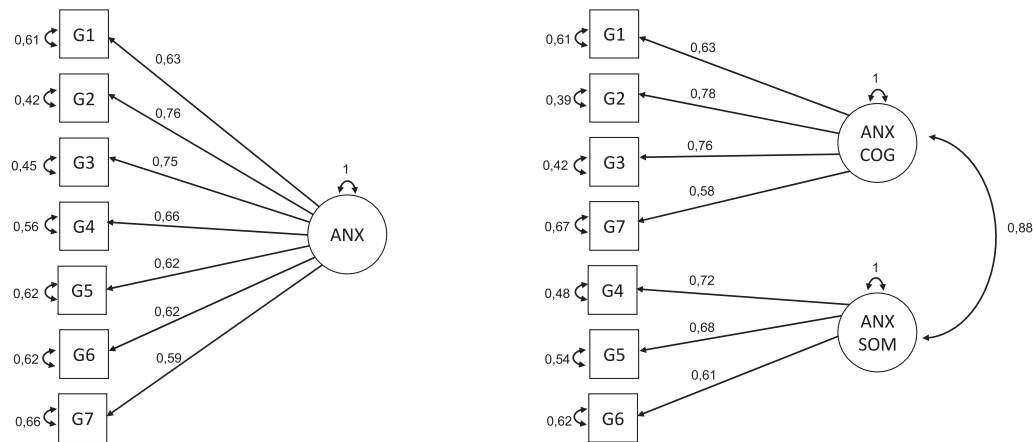


Fig. 2. Standardized factor loadings and error variances for 1- and 2-factor models of the GAD-7.

0.08) (see Table 5).

4. Discussion

Since depression and anxiety are the leading causes of disability, morbidity, and mortality among young women in Africa (Materu et al., 2020), there is a need for easy-to-use, short, and valid scales the identification of psychological distress symptoms in this population. This study aimed to examine the internal consistency reliability and construct validity of the PHQ-9 and GAD-7 to determine whether these measures were psychometrically reliable and valid scales for use among 18- to 28-year-old women living in Soweto. The study's use of Cronbach's alpha, Mokken scale analysis, and CFA provided a robust exploratory assessment of reliability and validity.

The internal consistency the PHQ-9 and GAD-7 was high, as demonstrated by a Cronbach's α of 0.81 and 0.84 respectively. This was additionally supported by positive inter-item correlations (PHQ-9 average $r = 0.33$; GAD-7 average $r = 0.44$), and item-by-scale correlations (PHQ-9 average $r = 0.63$; GAD-7 average $r = 0.72$). According to the PHQ-9 and the GAD-7, 14 % of young women in this study met the criteria for probable depression, 12 % showed symptoms of probable anxiety, and 9 % met the criteria for comorbid anxiety and depression. These prevalence rates align with a large-scale nationally representative

weighted survey conducted in 2021 (Craig et al., 2022), as well as other studies on young South African women in this setting reporting symptoms of probable depression and anxiety (Draper et al., 2022a, 2022b; Draper et al., 2023; Redinger et al., 2020), confirming that the mental health needs of young women should be prioritized as a relatively large number of young South African women are at risk for developing mental health conditions.

The proportion of participants identified at risk for depression (31 %), anxiety (29 %) and comorbid anxiety and depression (17 %) did not display sufficient depressive and/or anxiety symptomology to warrant a diagnosis. They may be unlikely to seek out treatment, but early detection and intervention are still recommended to reduce the more severe progression of these NCDs. Of particular concern were the 15 % of participants who reported agreement on P9 (*Thoughts that you would be better off dead or of hurting yourself in some way*) as this indicated young women who were at risk for suicidality and/or self-harm. Since only 14 % were classified with probable depression, this suggests that a number of participants answered "yes" to this but did not score ≥ 10 on the PHQ-9. As suggested in other studies (Howard et al., 2011; Redinger et al., 2020), it may be useful to explore whether using this one item of the PHQ-9 can help identify those at risk of harming themselves to provide early intervention, even if they do not meet the full criteria of probable depression.

With regards to factorial validity, CFA of the PHQ-9 and GAD-7 as separate scales showed a reasonable fit to a 1-factor model which supports the presence of a dominant general factor (depression or anxiety); as well as provided support for the 2-factor model which represented somatic and cognitive subtypes of depression and anxiety. The factor loadings on all models exceeded the conventional value of 0.30 and this was supported by the corrected item-total correlation values which indicated that all the items contributed meaningfully to each scale. The CFA supports the results of the Mokken analysis in demonstrating that the 1-factor models and 2-factor models provide reasonable fit and scale statistics and can both be used in this population of young South African women. However, the 2-factor models may offer additional clinical value for this population by allowing clinicians to specifically identify somatic and cognitive symptomology in young women at risk for depression and anxiety, which can be used to support diagnosis and treatment planning.

Support for the 2-factor model with somatic and cognitive subtypes has been found in other studies (Boothroyd et al., 2018; De Man et al., 2021), demonstrating that two (or more) clinically distinct profiles of symptoms may be found in some populations, which can include young South African women. Whilst the comorbidity of depression and anxiety has been established for young women in Africa, there are distinct differences between depression and anxiety and hence should be viewed as separate, yet related conditions. This has been supported by neurobiological findings (Lamela et al., 2020) which show that some individuals experiencing depression and/or anxiety may experience a combination of high levels of both subtypes of symptoms or experience higher levels of either somatic or cognitive symptoms. Other studies have found that depression and anxiety are typically expressed and experienced somatically in many African contexts, where symptoms of depression and anxiety are described in behavioural terms, rather than cognitive terms (Almedom et al., 2003; Bunting and Wessels, 1991; Uwakwe, 2003). “The spirit is down” (umoya uphansi), “the body is down” (umzimba uphansi), “the heart is sore” (inhliziyi ibuhlungu), and “not feeling well” (ukuphatheka kabi) are some of the ways South Africans have been found to describe their mental health symptoms (Sweetland et al., 2014). When South Africans complete self-reported assessment measures of depression and anxiety, it is possible that they may be more likely to self-report experiencing somatic indicators of mental health conditions, such as “*Trouble falling or staying asleep or sleeping too much*” (P3), “*Feeling tired or having little energy*” (P4) and “*Poor appetite or overeating*” (P5), and less likely to self-report on experiencing cognitive indicators of depression, such as “*Trouble concentrating on things, such as reading the newspaper or watching television*” (P7) “*Feeling down, depressed or hopeless*” (P2). This study confirmed that certain cognitive items were weaker when administered to a population of young South African women, which requires further investigation as this may suggest somatic indicators of depression and anxiety are more clinically relevant for diagnosis and treatment in this population.

Due to limited resources, many young South African women may not access the mental health treatment that they require, contributing to a higher risk of disability, morbidity, and mortality (Materu et al., 2020). Women who have limited access to social capital and economic opportunities due to their lowered social status are less likely to or able to access mental health interventions (Onuh et al., 2021). South African women are also exposed to social and environmental challenges that have been shown to impact negatively on their mental health, such as social injustice, gender-based violence and other traumatic events, and limited educational and employment opportunities (Draper et al., 2019; Redinger et al., 2020; Ware et al., 2021). One study established that one-quarter of young women living in Soweto were classified as socially vulnerable (Ware et al., 2021). Marginalized groups, including women and children, are considered to have the least power in society and tend to suffer the greatest psychological distress due to their experiences of disempowerment and subjugation (Rosenfield, 2012). Issues of gender inequality and strict gender norms that disadvantage women are often

more marked in LMICs, which places women at greater risk of social and physical vulnerabilities that contribute to poor mental health outcomes (Onuh et al., 2021; Walker, 2018; Ware et al., 2021). Whilst the use of suitable mental health instruments to identify probable depression and anxiety is a primary concern to ensure early detection, appropriate referral for treatment, and reduce the burden of depression and anxiety, how contextual and gender issues may complicate the subsequent development and implementation of such interventions must also be addressed.

The categorical classification of mental health conditions informs public health reporting and policymaking globally, as well as epidemiological research on mental health risk and protective factors. The prevalence of mental health conditions is used to guide the prevention, detection, and treatment of mental health conditions within a public health paradigm. Since “you can't manage what you do not measure” (Hayes et al., 2023) in global health, assessment of the prevalence of mental health conditions must be valid and reliable. A major challenge in limited resource contexts is the lack of appropriately developed and validated scales and the lack of specialists both in research and practice (Marlow et al., 2022). For example, the DSM criteria is used by South African psychiatrists and psychologists as there are limited reliable and valid diagnostic scales have been specially developed for use (Hook and Eagle, 2002). Furthermore, the lack of accessible mental health care and the underfunding of mental health services in South Africa means that mental health conditions often remain undiagnosed and untreated as patients cannot access the help they need (Sorsdahl et al., 2010). This contributes to the underreporting of mental health burden at a population level, which in turn means that service delivery of mental health services remains underdeveloped and underfunded (Docrat et al., 2019). South Africa's latest National Mental Health Policy Framework and Strategic Plan (2023–2030) highlights the importance of implementing clinical protocols for mental health assessment. Assessing the reliability and validity of mental health scales is a fundamental step to ensure the mental health burden is accurately reflected, so that public health policymakers make the right decisions about mental health service provision.

There remains a lack of consensus on whether mental health conditions are universal or whether symptoms are relative to culture (Canino and Alegria, 2008; Alarcón, 2009). Considering that the PHQ-9 and GAD-7 were developed and evaluated in Europe and North America, when these scales are used in other contexts, such as South Africa, questions are raised about how culture and contextual factors may affect the assessment and treatment of mental health conditions. The development of classification systems, such as the DSM, aims to provide diagnostic consensus across different contexts and cultures. Since the PHQ-9 and GAD-7 are scales based on the DSM criteria for depression and anxiety (APA, 2013), the universalist assumption that these scales can be used cross-culturally informed the decision to assess their appropriateness in a South African context. The underlying assumption is that mental health conditions can be identified as core symptoms that cluster into universal syndromal patterns that can be observed across cultures (Canino et al., 1997). Whilst the same mental health condition may manifest differently in different cultures and at different developmental stages, the underlying psychopathology, such as depression and anxiety, can be assessed cross-culturally using the same set of criteria and symptom clusters. Whether a diagnostic instrument based on a Western diagnostic system is culturally equivalent in a different culture from that in which the instrument was developed must be explored further; however, this was beyond the scope of this study. The lack of biological markers, imprecise scales, and the absence of a gold standard for validating mental health conditions further complicate mental health assessment and treatment practices (Canino and Alegria, 2008; Alarcón, 2009). Whilst this study shows that the PHQ-9 and GAD-7 show potential for reliably identifying probable depression and anxiety (with cognitive or somatic subtypes), there remains a need to critically understand the cultural issues concerning depression and anxiety of young

women in Soweto that may influence how these scales are interpreted and used to inform diagnosis and treatment in this population group.

4.1. Strengths and limitations

The results should be considered in relation to the following strengths and limitations. Whilst the PHQ-9 is a reliable unidimensional measure for depression (1-factor model) and the GAD-7 a reliable unidimensional measure for anxiety (1-factor model) in this population, further investigation into the cognitive or somatic subtypes (2-factor models) should consider how internal consistency and scalability were generally weaker for specific cognitive items from these measures, especially P8 (*Moving or speaking so slowly that other people could have noticed. Or the opposite*) and P7 (*Trouble concentrating on things, such as reading the newspaper or watching television*). These two items specifically refer to cognitive symptoms. The distinction between somatic and cognitive subtypes should be clinically considered for prognosis and treatment for depression and anxiety (Lamela et al., 2020). It may be beneficial to include additional analyses such as exploratory factor analysis (EFA) to complement CFA, as well as using alternative robust estimation methods if non-normality persists, as this could provide a more nuanced understanding of the factor structures.

The PHQ-9 and GAD-7 are self-report measures, so it is possible that social desirability, sampling, and response biases also influenced the results. Since this study took place within urban Soweto, South Africa, with a sample of young African women who shared competency in one language (English), the generalizability to other populations of African women, particularly in more rural areas, may be limited. To enhance generalizability, further psychometric studies with samples of women throughout South Africa are required to draw conclusions about the use of the PHQ-9 and the GAD-7 in this age group, as well as adaptation and translation work of measures with non-English speaking samples of African women. Future research could also explore the scales' validity across different regions and among various demographic groups within Africa to evaluate the scales' applicability across different contexts with boarder and more diverse samples. The potential influence of cultural and linguistic factors on the scales' effectiveness is also an important consideration. Research could include a cultural adaptation process for the scales and evaluate how such adaptations affect their validity and reliability.

This study highlights that whilst the PHQ-9 and GAD-7 are valid for identifying those in the population who are potentially at-risk for depression and anxiety, they should not replace comprehensive diagnostic assessment and clinical diagnosis. This distinction is crucial for the proper use of these scales in clinical practice. Whilst the PHQ-9 and GAD-7 have been recommended for use in settings with poor mental health service provision and a lack of specialized staff to identify the risk of mental health conditions (Mughal et al., 2020), diagnosis should include a multimodal assessment process as this is the most effective way to assess and diagnose any mental health disorder as it accounts for the limitations of each measure included. The use of measures like the PHQ-9 and GAD-7 could allow for screening so that patients at risk can be referred for an appropriate and comprehensive diagnostic assessment. Due to shortages of mental health specialists in South Africa and many other LMICs, comprehensive, high-quality diagnostic assessments remain a challenge (Marlow et al., 2022). Future studies would need to examine how multimodal diagnostic assessment could be feasible in limited resource contexts to ensure the delivery of appropriate mental health diagnosis and treatment. Future research could also explore the diagnostic accuracy of these scales by comparing them against clinical diagnoses and assessing their ability to predict treatment outcomes.

Whilst this study significantly contributes to the understanding of the PHQ-9 and GAD-7 scales in a specific context, the assessment of the psychometric properties of mental health measures is considered an iterative process that requires validation across diverse samples, as well as corroboration by alternative psychometric methods. Validation is an

extensive process that should include local experts and mental health advocacy groups in cross-cultural contexts (Marlow et al., 2022). Whilst this study targeted two measures for inclusion in testing models for robustness for a specific population of young women in Soweto, it is imperative that measures are continually evaluated to ensure their appropriateness, especially across diverse settings and populations. The PHQ-9 and the GAD-7 were found to show potential for identifying those at risk for depression and anxiety in this sample of young Sowetan women. In terms of reliability, the measures were established to have acceptable internal consistency. The reliability of the PHQ-9 and the GAD-7 can be further investigated in this population by examining whether participants' scores are stable across items over the course of the trial (reproducibility). This study provided evidence for the construct validity of the PHQ-9 and the GAD-7; however, future studies should explore how these measures compare to previously validated instruments that screen for depression and anxiety (concurrent validity), as well as whether the PHQ-9 and the GAD-7 scores differ according to known groups in predictable ways (known-groups validity tests). Including a gold standard measure of depression and anxiety will provide the necessary support for the diagnostic properties (sensitivity and specificity) of the PHQ-9 and the GAD-7 for depression and anxiety respectively. Addressing the noted limitations and expanding the research scope can further solidify the scales' utility and reliability in diverse settings.

5. Conclusion

This study evaluated the psychometric properties of the PHQ-9 and GAD-7 among young women in Soweto, South Africa, revealing that both scales demonstrated high internal consistency and construct validity in this population. The scales' Cronbach's alpha values, positive item correlations, and factorial validity supported their reliability and appropriateness for identifying psychological distress among young South African women. The observed prevalence rates of probable depression and anxiety aligned with existing literature, underscoring the critical need for targeted mental health interventions in this demographic.

Whilst the PHQ-9 and GAD-7 effectively can be used to identify general patterns of depression and anxiety, there are concerns about the lower reliability of cognitive items, suggesting that somatic symptoms may be more salient in this cultural context. This points to the need for further research to adapt these scales to better capture culturally specific manifestations of mental health conditions. Additionally, given the limited access to mental health resources in South Africa, these scales can facilitate early identification but should be complemented by comprehensive diagnostic assessments.

Future research should focus on validating these measures in diverse African contexts, considering cultural and linguistic variations. Exploring multimodal diagnostic approaches and comparing these scales with other validated instruments could enhance their utility and accuracy. Continuous evaluation and adaptation are essential to ensure that mental health scales remain effective and relevant across different populations and settings. Ultimately, improving the precision of mental health assessments will aid in the development of more effective interventions, helping to address the significant mental health needs of African young women.

Funding

Open access funding provided by University of the Witwatersrand. This work was supported by the South African Medical Research Council, and the Canadian Institutes of Health Research. Whilst the funders provided initial input on the design of the study, the funders have not been involved in the collection, analysis, and interpretation of the data for this study.

CRediT authorship contribution statement

Claire Hart: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Catherine E. Draper:** Writing – review & editing, Supervision, Conceptualization. **Larske M. Soepnel:** Writing – review & editing, Software, Conceptualization. **Motlatso Godongwana:** Writing – review & editing, Conceptualization. **Khuthala Mabetha:** Writing – review & editing, Supervision, Conceptualization. **Lukhanyo H. Nyati:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Simone H. Crouch:** Writing – review & editing, Supervision, Conceptualization. **Shane A. Norris:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

The authors acknowledge the South African Medical Research Council and the Canadian Institutes of Health Research for funding *Bukhali*. We extend our gratitude to the women who are participating in the study and the staff who are implementing this trial. We thank the WHO, the HeLTI PI Research Committee, and HeLTI Office for their contributions and support to *Bukhali*.

References

- Alarcón, R.D., 2009. Culture, cultural factors and psychiatric diagnosis: review and projections. *World Psychiatry* 8 (3), 131–139. <https://doi.org/10.1002/j.2051-5545.2009.tb00233.x>.
- Almedom, A.M., Tesfamichael, B., Yacob, A., et al., 2003. Maternal psychosocial well-being in Eritrea: application of participatory methods and tools of investigation and analysis in complex emergency settings. *Bull. World Health Organ.* 81 (5), 360–366. <https://doi.org/10.1590/S0042-96862003000500010>.
- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, 5th ed. American Psychiatric Publishing.
- Beard, C., Björgvinsson, T., 2014. Beyond Generalized Anxiety Disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *J. Anxiety Disord.* 28 (6), 547–552. <https://doi.org/10.1016/j.janxdis.2014.06.002>.
- Boothroyd, L., Dagnan, D., Muncer, S., 2018. Psychometric analysis of the Generalized Anxiety Disorder Scale and the Patient Health Questionnaire using Mokken scaling and confirmatory factor analysis. *Health Primary Care* 2. <https://doi.org/10.15761/HPC.1000145>.
- Buntting, B.G., Wessels, W.H., 1991. The “present state examination” and the structured clinical interview in Zulu. *S. Afr. Med. J.* 79 (2), 96–98.
- Canino, G., Alegría, M., 2008. Psychiatric diagnosis - is it universal or relative to culture? *J. Child Psychol. Psychiatry* 49 (3), 237–250. <https://doi.org/10.1111/j.1469-7610.2007.01854.x>.
- Canino, G., Lewis-Fernandez, R., Bravo, M., 1997. Methodological challenges in cross-cultural mental health research. *Transcult. Psychiatry*. <https://doi.org/10.1177/136346159703400201>.
- Clark, L.A., Watson, D., 1995. Constructing validity: basic issues in objective scale development. *Psychol. Assess.* 7 (3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>.
- Craig, A., Rochat, T., Naicker, S.N., Mapanga, W., Mtintsilana, A., Dlamini, S.N., Ware, L. J., Du Toit, J., Draper, C.E., Richter, L., Norris, S.A., 2022. The prevalence of probable depression and probable anxiety, and associations with adverse childhood experiences and socio-demographics. *Front. Public Health* 10, 986531. <https://doi.org/10.3389/fpubh.2022.986531>.
- De Man, J., Absetz, P., Sathish, T., Desloge, A., Haregu, T., Oldenburg, B., Johnson, L.C. M., Thankappan, K.R., William, E.D., 2021. Are the PHQ-9 and GAD-7 suitable for use in India? A psychometric analysis. *Front. Psychol.* 12. <https://doi.org/10.3389/fpsyg.2021.676398>.
- Docrat, S., Besada, D., Cleary, S., Daviaud, E., Lund, C., 2019. Mental health system costs, resources and constraints in South Africa: a national survey. *Health Policy Plan.* 34 (9), 706–719. <https://doi.org/10.1093/heapol/czz085>.
- Doi, S., Ito, M., Takebayashi, Y., Muramatsu, K., Horikoshi, M., 2018. Factorial validity and invariance of the Patient Health Questionnaire (PHQ)-9 among clinical and non-clinical populations. *PLoS One* 13 (7), e0199235. <https://doi.org/10.1371/journal.pone.0199235>.
- Draper, C.E., Bosire, E., Prioreschi, A., Ware, L.J., Cohen, E., Lye, S.J., et al., 2019. Urban young women's preferences for intervention strategies to promote physical and mental health preconception: a healthy life trajectories initiative (HeLTI). *Prev. Med.* Rep. 14, 100846. <https://doi.org/10.1016/j.pmedr.2019.100846>.
- Draper, C.E., Cook, C.J., Redinger, S., Rochat, T.J., Prioreschi, A., Rae, D.E., Ware, L.J., Lye, S.J., Norris, S.A., 2022a. Cross-sectional associations between mental health indicators and social vulnerability, with physical activity, sedentary behaviour and sleep in urban African young women. *Int. J. Behav. Nutr. Phys. Act.* 19 (82). <https://doi.org/10.1186/s12966-022-01325-w>.
- Draper, C.E., Mabena, G., Motlathledi, M., Thwala, N., Lawrence, W., Weller, S., Klingberg, S., Ware, L.J., Lye, S.J., Norris, S.A., 2022b. Implementation of Healthy Conversation Skills to support behaviour change in the *Bukhali* trial in Soweto, South Africa: a process evaluation. *SSM—Ment. Health* 2, 100132. <https://doi.org/10.1016/j.ssmmh.2022.100132>.
- Draper, C.E., Thwala, N., Slemming, W., Lye, S.J., Norris, S.A., 2023. Development, implementation, and process evaluation of *Bukhali*: an intervention from preconception to early childhood. *Glob. Implement. Res. Appl.* <https://doi.org/10.1007/s43477-023-00073-8>.
- Hayes, J., Carvajal-Velez, L., Hijazi, Z., Ahs, J.W., Doraiswamy, P.M., El Azzouzi, F.A., Fox, C., Herrman, H., Gornitzka, C.P., Staglin, B., Wolpert, M., 2023. You can't manage what you do not measure - why adolescent mental health monitoring matters. *J. Adolesc. Health* 72 (1S), S7–S8. <https://doi.org/10.1016/j.jadohealth.2021.04.024>.
- Hook, D., Eagle, G., 2002. *Psychopathology and Social Prejudice*. University of Cape Town Press, Cape Town.
- Howard, L.M., Flach, C., Mehay, A., Sharp, D., Tylee, A., 2011. The prevalence of suicidal ideation identified by the Edinburgh Postnatal Depression Scale in postpartum women in primary care: findings from the RESPOND trial. *BMC Pregnancy Childbirth* 11 (1), 57. <https://doi.org/10.1186/1471-2393-11-57>.
- Jewkes, R.K., Dunkle, K., Nduna, M., Jama, P.N., Puren, A., 2010. Associations between childhood adversity and depression, substance abuse and HIV and HSV2 incident infections in rural South African youth. *Child Abuse Negl.* 34 (11), 833–841. <https://doi.org/10.1016/j.chiabu.2010.05.002>.
- Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16 (9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>.
- Kroenke, K., Spitzer, R.L., Williams, J.B., Monahan, P.O., Löwe, B., 2007. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann. Intern. Med.* 146 (5), 317–325. <https://doi.org/10.7326/0003-4819-146-5-200703060-00004>.
- Lamela, D., Soreira, C., Matos, P., Morais, A., 2020. Systematic review of the factor structure and measurement invariance of the Patient Health Questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings. *J. Affect. Disord.* 276, 220–233. <https://doi.org/10.1016/j.jad.2020.06.066>.
- Marlow, M., Skeen, S., Grieve, C.M., Carvajal, L., Åhs, J.W., Kohrt, B.A., Requejo, J., Stewart, J., Henry, J., Goldstone, D., Kara, T., Tomlinson, M., 2022. Detecting depression and anxiety among adolescents in South Africa: validity of the isiXhosa Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7. *J. Adolesc. Health*. <https://doi.org/10.1016/j.jadohealth.2022.09.013>.
- Materu, J., Kuringe, E., Nyato, D., Galishi, A., Mwanamsangu, A., Katebalila, M., Shao, A., Changalucha, J., Nnko, S., Wambura, M., 2020. The psychometric properties of PHQ-4 anxiety and depression screening scale among out of school adolescent girls and young women in Tanzania: a cross-sectional study. *BMC Psychiatry* 20 (1), 321. <https://doi.org/10.1186/s12888-020-02735-5>.
- Mkhwanazi, S., Gibbs, A., 2021. Risk factors for Generalized Anxiety Disorder among young women and men in informal settlements in South Africa: a cross-sectional study. *SSM - Ment. Health* 1, 100010. <https://doi.org/10.1016/j.ssmmh.2021.100010>.
- Mokken, R.J., 1971. *A Theory and Procedure of Scale Analysis*. De Gruyter.
- Monahan, P.O., Shacham, E., Reece, M., et al., 2009. Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. *J. Gen. Intern. Med.* 24 (2), 189–197. <https://doi.org/10.1007/s11606-008-0846-z>.
- Mughal, A.Y., Devadas, J., Ardman, E., Levis, B., Go, V.F., Gaynes, B.N., 2020. A systematic review of validated screening tools for anxiety disorders and PTSD in low to middle income countries. *BMC Psychiatry* 20, 338. <https://doi.org/10.1186/s12888-020-02753-3>.
- Norris, S.A., Draper, C.E., Smuts, C., Prioreschi, A., Ware, L.J., Dennis, C.L., Awadalla, P., Bassani, D.G., Bhutta, Z.A., Briollais, L., Cameron, B., Chirwa, T., Fallon, B., Gray, C., Hamilton, J., Jamison, J., Jaspán, H., Jenkins, J.M., Kahn, K., Lye, S.J., 2022. Building knowledge, optimising physical and mental health, and setting up healthier life trajectories in South African women (*Bukhali*): a preconception randomised control trial part of the Healthy Life Trajectories Initiative (HeLTI). *BMJ Open* 12, e059914. <https://doi.org/10.1136/bmjopen-2021-059914>.
- Onuh, J.C., Mbah, P.O., Ajaero, C.K., Orjiakor, C.T., Igboeli, E.E., Ayogu, C.K., 2021. Rural-urban appraisal of the prevalence and factors of depression status in South Africa. *J. Affect. Disord. Rep.* 4. <https://doi.org/10.1016/j.jadr.2021.100082>.
- Osborn, T.L., Venturo-Conerly, K.E., Wasil, A.R., et al., 2020. Depression and anxiety symptoms, social support, and demographic factors among Kenyan high school students. *J. Child Fam. Stud.* 29, 1432–1443. <https://doi.org/10.1007/s10826-019-01646-8>.
- Redinger, R., Pearson, R.M., Houle, B., Norris, S.A., Jean Rochat, T., 2020. Antenatal depression and anxiety across pregnancy in urban South Africa. *J. Affect. Disord.* 277, 296–305. <https://doi.org/10.1016/j.jad.2020.08.010>.
- Rosenfield, S., 2012. Triple jeopardy? Mental health at the intersection of gender, race, and class. *Soc. Sci. Med.* 74 (11), 1791–1801. <https://doi.org/10.1016/j.socscimed.2011.11.010>.

- Rutter, L.A., Brown, T.A., 2017. Psychometric properties of the Generalized Anxiety Disorder scale-7 (GAD-7) in outpatients with anxiety and mood disorders. *J. Psychopathol. Behav. Assess.* 39, 140–146. <https://doi.org/10.1007/s10862-016-9571-9>.
- Santomauro, et al., 2021. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the Covid-19 pandemic. *Lancet* 398 (10312), 1700–1712. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7).
- Sorsdahl, K., Flisher, A.J., Wilson, Z., Stein, D.J., 2010. Explanatory models of mental disorders and treatment practices among traditional healers in Mpumalanga, South Africa. *Afr. J. Psychiatry* 13, 284–290.
- Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* 166 (10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>.
- Stein, A., Pearson, R.M., Goodman, S.H., Rapa, E., Rahman, A., McCallum, M., Howard, L.M., Pariente, C.M., 2014. Effects of perinatal mental disorders on the fetus and child. *Lancet* 384 (9956), 1800–1819. [https://doi.org/10.1016/S0140-6736\(14\)61277-0](https://doi.org/10.1016/S0140-6736(14)61277-0).
- Streiner, D.L., 2003. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J. Pers. Assess.* 80 (1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18.
- Sweetland, A.C., Belkin, G.S., Verdeli, H., 2014. Measuring depression and anxiety in sub-Saharan Africa. *Depress. Anxiety* 31 (3), 223–232. <https://doi.org/10.1002/da.22142>.
- Uwakwe, R., 2003. Affective (depressive) morbidity in puerperal Nigerian women: validation of the Edinburgh Postnatal Depression Scale. *Acta Psychiatr. Scand.* 107, 251–259. <https://doi.org/10.1034/j.1600-0447.2003.02477.x>.
- Vittengl, J.R., Clark, L.A., Smits, J.A.J., Thase, M.E., Jarrett, R.B., 2019. Do comorbid social and other anxiety disorders predict outcomes during and after cognitive therapy for depression? *J. Affect. Disord.* 242, 150–158. <https://doi.org/10.1016/j.jad.2018.08.053>.
- Walker, M., 2018. Aspirations and equality in higher education: gender in a South African university. *Camb. J. Educ.* 48, 123–139. <https://doi.org/10.1080/0305764X.2016.1254159>.
- Ware, L.J., Kim, A.W., Prioreschi, A., Nyati, L.H., Taljaard, W., Draper, C.E., et al., 2021. Social vulnerability, parity and food insecurity in urban south African young women: the healthy life trajectories initiative (HeLTI) study. *J. Public Health Policy* 42 (3), 373–389. <https://doi.org/10.1057/s41271-021-00289-8>.
- Weiss, S.J., Simeonova, D.I., Kimmel, M.C., Battle, C.L., Maki, P.M., Flynn, H.A., 2016. Anxiety and physical health problems increase the odds of women having more severe symptoms of depression. *Arch. Womens Ment. Health* 19. <https://doi.org/10.1007/s00737-015-0575-3>.