

RESEARCH ARTICLE

Open Access



# 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

Jacqueline Materu<sup>1\*</sup> , Evodius Kuringe<sup>1</sup>, Daniel Nyato<sup>1</sup>, Anthony Galishi<sup>2</sup>, Amasha Mwanamsangu<sup>2</sup>, Maligo Katebalila<sup>2</sup>, Amani Shao<sup>1</sup>, John Changalucha<sup>1</sup>, Soori Nnko<sup>1</sup> and Mwita Wambura<sup>1</sup>

## Abstract

**Background:** Literature suggests that most mental disorders have their onset in childhood and adolescence, but go undiagnosed until adulthood. Shorter versions of the screening tools such as the Patient Health Questionnaire with four items (PHQ-4) may help to improve screening coverage. This study assessed the psychometric properties of the PHQ-4 in screening for core symptoms of depression and anxiety among out of school adolescent girls and young women (AGYW).

**Methods:** This is a cross-sectional analysis of data from a cluster randomized controlled trial conducted among AGYW between June and July 2018 in North-West Tanzania. Two thousand four hundred twenty-six out-of-school AGYW aged 15 to 23 years were included. Data were collected on tablets using audio computer-assisted self-interviews (ACASI). Cronbach's  $\alpha$  was used to measure the reliability of the PHQ-4 while confirmatory factor analysis (CFA) and principal components analysis (PCA) were used for construct validity assessment. In CFA, three criteria were used to assess how well the model fits the data: Standardized Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA) and 90% confidence interval for RMSEA.

**Results:** Of the 2426 participants, 33.8 and 35.5% screened positive for core symptoms of anxiety ( $GAD-2 \geq 3$ ) and depression ( $PHQ-2 \geq 3$ ), respectively. Cronbach's  $\alpha$  of the PHQ-4 was 0.81. Both items-correlation and corrected items-correlation of the PHQ-4 had total correlations above 0.5 ( $p < 0.01$ ). CFA showed that all items loaded significantly onto the single factor, and loadings were strong, ranging from 0.67 to 0.77 ( $p < 0.01$ ). CFA indicates that the PHQ-4 scale stand for a unidimensional construct with good model fit ( $CFI = 0.995$ ,  $SRMR = 0.013$ ,  $RMSEA = 0.054$  and 90% CI for RMSEA (0.031–0.079)). PCA confirmed two distinct components;  $GAD-2$  (anxiety) and  $PHQ-2$  (depression). Those who reported having suicidal thoughts and social function problems had significantly higher scores on  $PHQ-2$ ,  $GAD-2$ , and  $PHQ-4$  screening items ( $p < 0.01$ ).

**Conclusions:** The findings suggest that the PHQ-4 scale can reliably and validly screen for core symptoms of depression and anxiety among out of school AGYW. This tool is short and easy to administer. Thus, the PHQ-4 scale can be very useful in screening for anxiety and depression symptoms in the community, primary health facilities, research and programmatic settings.

**Keywords:** Anxiety, Depression, PHQ-4, Reliability, Validity, Adolescent girls and young women, Tanzania

\* Correspondence: [jacqueline.materu1989@gmail.com](mailto:jacqueline.materu1989@gmail.com)

<sup>1</sup>Department of Sexual and Reproductive Health, National Institute for Medical Research, Mwanza Centre, P.O Box 1462, Mwanza, Tanzania  
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Depression and anxiety are the leading causes of disability, morbidity, and mortality among adolescent girls and young women (AGYW) aged 15 to 29 years in Africa [1]. Evidence increasingly shows varied prevalence across different populations including AGYW and people living with HIV (PLHIV) [2, 3]. For instance, a study conducted in Kenya and Zambia using the four-item Patient Health Questionnaire (PHQ-4) reported the prevalence of moderate and severe symptoms of anxiety and depression of 3.7 and 9.4% respectively, among AGYW [2]. A study conducted in Tanzania reported a prevalence of depression of 27% among children and adolescents living with HIV [3].

Literature suggests that most mental disorders go undiagnosed until adulthood, even though their onset is around childhood and adolescence [4–6]. Factors related to access to mental health services contribute to this gap. These include individual factors such as lack of awareness, societal factors, including myths regarding mental health disorders and systemic factors, for example, the centralization of services and inadequately qualified staff [7, 8]. Consequently, AGYW with anxiety and depression continue to experience functional impairment, academic difficulties, suicidal ideation and attempts, and further risk of the futile cycle of mental illnesses [9–12]. Moreover, children born by mothers with these conditions may also suffer suboptimal cognitive development [13].

Several recommendations have been put forward in the literature on how to improve access and utilization of mental health services. These recommendations include decentralization and task shifting in the provision of mental health services, to maximize uptake of services both at lower-level facilities and the community [8, 14]. To be able to realize this, screening scales must be made available to enable identification of probable mental health disorder patients and provide referral to health facilities for diagnosis. Scales such as Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorders-7 (GAD-7) have been used for depression and anxiety, respectively. However, to improve efficiency during mass screening or in busy outpatient clinics, shorter versions of the scales have been researched and successfully validated among clinic-based and general populations [15, 16]. One of the shorter versions of the screening scales is the Patient Health Questionnaire-4 (PHQ-4) [9]. This is a four-item screening tool for both depression and anxiety [9]. This tool was validated in the general adult population in Germany [16] and among college students in the United States [17]. A study conducted among college students in the United States demonstrated the PHQ-4 scale to be an effective tool for screening depression and anxiety, with the area

under the curve (AUC) of 0.835 and 0.787 for depression and anxiety respectively [17]. While validation studies have been conducted in other settings [16, 17], findings from these studies cannot be generalized to Tanzania context, given the cultural and study population differences (out of school AGYW in Tanzania versus college students in the USA [17] and general adult population in Germany [16]). Besides, there are also differences in survey administration methods where the present study uses audio computer-assisted self-interviews (ACASI) while previous studies used interviewer administered techniques [16, 17]. Therefore, this study aims to explore the reliability and validity of the PHQ-4 in a large out-of-school AGYW population in Tanzania using ACASI data collection technique.

## Methods

This study uses data from a cluster randomized controlled trial, CARE study, conducted in Kahama, Ushetu and Msalala districts. CARE study was approved by the Medical Research Coordination Committee of the National Institute for Medical Research (NIMR/HQ/R.8a/Vol.IX/2287) and the Johns Hopkins University IRB (00007976). CARE study is also registered at ClinicalTrials.gov, number NCT03597243. Details on CARE study are documented elsewhere [18].

## Study population

Eligible for the study were AGYW aged 15 to 23 years who were out of school, residents of the village selected for study participation and who have completed 10 h of social and behavior change communication (SBCC) sessions. Besides, AGYW were eligible if they provided voluntary informed consent to take part in the study. The current study utilizes data from the first follow-up round of CARE study ( $n = 2426$ ). Participants were enrolled from 30 villages from June to July 2018. The authors used second-round data because the data consists of additional information on mental health disorders, including the effect of depression and anxiety symptoms on daily activities and social relationships, and suicidal ideation.

## Procedures

During SBCC sessions, potential participants were informed about the dates and locations of CARE study recruitment. Thus, the potential participants were able to reach the recruitment site for pre-screening consent and eligibility screening at baseline until the sample required was reached. During follow up rounds, participants were also informed of data collection dates and locations in their SBCC groups.

Within the CARE study, the study participants were interviewed using audio computer-assisted self-interviews

(ACASI). ACASI is an electronic self-administered questionnaire, where participants listen privately to pre-recorded interview questions using a tablet connected to earphones [18, 19]. The participants could also simultaneously read each interview question, including the PHQ-4 items, on the tablet screen as the question was being read. The interviews were conducted in settings within the community, which ensured the privacy of study participants. Data collection was facilitated by research assistants who were trained on research ethics, interviewing skills, data management as well as study-specific procedures. The research assistants also received online training in good clinical practice (GCP).

### Instrument and measures

The ACASI questionnaire included questions on demographics and mental health symptoms (S1 Questionnaire). Mental health screening was conducted using the 4-item PHQ for depression and anxiety developed by Kroenke and colleagues [9].

The PHQ-4 is a concise tool composed of the PHQ-2 screening tool for depression and generalized anxiety disorders (GAD) screener (GAD-2) [9]. PHQ-2 collects self-reports of two core symptoms of depression according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) [9]. The depression questions were phrased as: “Over the *last 2 weeks*, how often have you been bothered by the following problems?”; ‘Feeling down, depressed, or hopeless’ and ‘Little interest or pleasure in doing things’ [9]. While GAD-2 collects two core symptoms of anxiety and was phrased as “Over the *last 2 weeks*, how often have you been bothered by the following problems?”; ‘Feeling nervous or anxious or on edge’ and ‘Not being able to stop or control worrying’ [9]. Participants responded using a 4-point Likert-type of options; ‘not at all’ =0, ‘several days but less than one week’ =1, ‘more than half the days’ =2, ‘nearly every day’ =3. Positive screening for anxiety was assigned for participants if the score of the two core symptoms of anxiety was greater than or equal to three (GAD-2  $\geq 3$ ) [9]. Similarly, for the two core symptoms of depression, scores of greater than or equal to three (PHQ-2  $\geq 3$ ) were assigned positive screening [9]. Lastly, the overall measure of anxiety and depression was graded as normal (0–2), mild (3–5), moderate (6–8), and severe (9–12) [9].

Two additional questions were also obtained from the Patient Health Questionnaire with nine items (PHQ-9) [20]. The PHQ-9 questions were worded as: ‘If you checked off *any* problems, how *difficult* have these problems made it for you to do your work, take care of things at home, or get along with other people?’ and ‘Have you had thoughts that you would be better off dead or of hurting yourself in some way in the last two

weeks’ [20]. All PHQ-4 data for the current study were collected as baseline assessments during the second round of data collection.

### Data management

Data from ACASI were collected using tablets and sent to the server located at the National Institute for Medical Research (NIMR) in Mwanza daily through a secure file transfer protocol. The data were extracted by the data manager for data quality checks. While the data collectors were still in the field, data queries were generated and sent to the field for resolution. The process continued until all the queries were solved, and then the final datasets were created (S1 Dataset).

### Data analysis

Data were managed and analysed using SAS 9.4 (SAS Institute Inc.; Cary, North Carolina) and STATA 14 (College Station, TX: StataCorp LP). Descriptive statistics (frequencies, means, and standard deviations) were used to describe the AGYW in terms of their demographic characteristics. An item analysis was conducted to assess score distributions, missing data patterns and floor/ceiling effects. Floor or ceiling effects were considered to be present if more than 15% of participants attained the lowest or highest possible scores, respectively [21, 22]. Analysis of score distribution was based on descriptive statistics for ordinal data. Item characteristics of the PHQ-4 items (item-means, item-intercorrelation and corrected item-total correlations) were explored.

For reliability, the internal consistency of the PHQ-4 was assessed. Cronbach’s alpha was used to measure internal consistency (i.e. the extent to which all items on a scale were inter-related or contribute positively towards measuring the same construct) [23–25]. Most of the time, the acceptable range of Cronbach’s alpha is a value of 0.70 or above [26]. However, it depends on how a measure is being applied and has been suggested that in research more stringent cut-offs (i.e. 0.80 or higher and 0.90 or higher) should be used [27]. Due to the ordinal nature of the PHQ-4 items, the estimation of the Cronbach’s alpha was based on the polychoric correlation matrix [28].

To explore whether the four items of the scale depicted a single dimension, confirmatory factor analysis (CFA) was used to assess model fit for a one-factor solution. Considering the ordinal nature of the data into account, the analysis was based on the two-stage estimation approach [29]. In stage one, the polychoric correlation was computed. Then, the model was analyzed with the appropriate weight matrix computed in stage one. This procedure ensures unbiased parameter estimates and standard error for ordinal data [29]. To get more accurate information regarding the goodness

of model fit, Standardized Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA) and 90% confidence interval for RMSEA were used [30]. CFI values greater than 0.94 indicate good fit [31], RMSEA values less than 0.09 indicate a fair or adequate error of approximation whereas values less than 0.055 indicate a small error [31–33] and RMSEA = 0 indicate exact model fit of approximation [34]. Likewise to the RMSEA, SRMR values less than 0.09 indicate fair or adequate fit, whereas SRMR values less than 0.055 indicate a good fit [31].

Principal components analysis (PCA) using a varimax-rotated component-matrix for the four screening items (two items each for GAD-2 and PHQ-2) were also used for construct validity assessment. In principal components analysis, the observed variables are transformed into a smaller number of artificial variables (i.e. principal components), which will account for most of the variance in the observed variables. The principal components can be used as a criterion variable in the consequent analysis [35, 36]. A varimax rotation (orthogonal rotation) method was used with the assumption that there is no correlation between components [9, 16, 37]. Furthermore, the association was checked between the PHQ-2, GAD-2, and PHQ-4 scores for respondents who had reported suicidal thoughts or social dysfunction versus those who had not reported suicidal thoughts or social dysfunction in the past 2 weeks.

## Results

### Participants' characteristics

A total of 2426 participants were interviewed. The mean age of participants was 20 years (standard deviation (SD) = 2.5), with 1393 (57%) being 20 to 23 years old. Over half of AGYW had a primary education (56%), 1224 married (51%) and 1550 had children (64%). The proportion of participants with positive screening for anxiety (GAD-2  $\geq$  3) was 33.8%, and depression (PHQ-2  $\geq$  3) was 35.5% (Table 1).

### Item characteristics and reliability

Descriptive statistics of the individual items for depression and anxiety screening sub-scale (i.e. PHQ-2 and GAD-2) and the total scale (i.e. PHQ-4) are shown in Table 2. The average anxiety score was 1.79, and the SD was 1.71 while for depression the average score was 1.96, and the SD was 1.66. The score distribution of the lowest response (i.e. not at all) ranged between 46.0 and 47.8% of the items in anxiety and between 35.0 and 48.2% for depression. No ceiling effects observed; all items showed floor effects. Both items-correlation and corrected items-correlation have total correlations above 0.5. All of these correlations were statistically significant at  $p < 0.01$ .

**Table 1** Demographic Characteristics of Participants

	n = 2426	%
	Mean ( $\pm$ SD)	
<b>Age (Years)</b>	19.7 (2.5)	
<b>Age (Years)</b>		
15–19	1033	42.6
20–23	1393	57.4
<b>Education level</b>		
No formal/Incomplete primary education	583	24.0
Primary school	1364	56.2
Complete/incomplete secondary school	479	19.7
<b>Marital status</b>		
Single	1039	42.8
Married (polygamous/monogamous/cohabiting)	1224	50.5
Formerly-married (separated/divorced/widowed)	163	6.7
<b>Family composition (Lives with)</b>		
Alone/ younger siblings/employer	51	2.1
Husband	1187	48.9
Parents/relative/ elder siblings	1188	49.0
<b>Ever had children</b>		
Yes	1550	63.9
No	876	36.1
<b>Living in TASAF households<sup>a</sup></b>		
Yes	212	8.7
No	2214	91.3
<b>Ever went to bed hungry due to lack of food<sup>b</sup></b>		
Yes	298	12.3
No	2128	87.7
<b>Anxiety symptoms (GAD-2 <math>\geq</math> 3)</b>	821	33.8
<b>Depression symptoms (PHQ-2 <math>\geq</math> 3)</b>	861	35.5

<sup>a</sup>Reported living in a household supported by a social welfare program; Tanzania Social Action Fund (TASAF); <sup>b</sup>Reported staying a day or going to bed hungry due to lack of food within the last 4 weeks

The Cronbach's alpha for PHQ-4, GAD-2, and PHQ-2 computed for internal-consistency (i.e. reliability), were  $\alpha = 0.81$ ,  $\alpha = 0.74$ , and  $\alpha = 0.65$  respectively. None of the items if deleted improves reliability (i.e. increases Cronbach's alpha) (Table 3).

### Construct validity

CFA results showed that all items loaded significantly onto the single factor, and loadings were strong, ranging from 0.67 to 0.77 ( $p < 0.01$ ). CFA indicate that the PHQ-4 scale stand for a unidimensional construct with good model fit (CFI = 0.995, SRMR = 0.013, RMSEA = 0.054 and 90% CI for RMSEA (0.031–0.079)) (Table 4). Principal components analysis (PCA) with the varimax-rotated component-matrix for the four screening items (i.e. two items each for

**Table 2** Item scores and scale correlations for study participants

Over the last two weeks, how often have you been bothered by the following problems	Mean (SD)	Score distribution %				<sup>a</sup> Item-total correlation	<sup>b</sup> Corrected item-total correlation
		0	1	2	3		
<b>Anxiety screener (GAD-2) sum score</b>	1.79 (1.71)						
1. Feeling nervous, anxious or on edge	0.92 (1.01)	<b>46.00</b>	25.06	19.54	9.40	0.76	0.64
2. Not being able to stop or control worrying	0.87 (0.98)	<b>47.82</b>	26.01	17.97	8.20	0.76	0.64
<b>Depression screener (PHQ-2) sum score</b>	1.96 (1.66)						
3. Little interest or pleasure in doing things	0.85 (0.97)	<b>48.15</b>	26.05	18.55	7.25	0.73	0.60
4. Feeling down, depressed or hopeless	1.11 (1.01)	<b>35.04</b>	30.87	22.42	11.67	0.74	0.59
<b>Total score (PHQ4)</b>	3.75 (2.98)						

Scale items scored 0, 1, 2, 3 based on responses provided (not at all, several days, more than half the days, and nearly every day).

<sup>a</sup>Correlation between the item and the total score from the scale (i.e. the item itself is included in the total); <sup>b</sup>Correlation between the item and the overall score from the scale (i.e. the item itself is not included in the total)

PHQ-2 and GAD-2) shows that the two components explained 82% of the total variance. Two anxiety screening items had the highest loading on component 1, and the two depression screening items had the highest loading on component 2 (Table 4). Besides, the association was checked between PHQ-2, GAD-2, and PHQ-4 scores for individuals who had reported social function problem and suicidal thoughts versus those who had not in the past 2 weeks (Table 5). Those who reported having suicidal thoughts and social function difficulties had significantly higher scores on PHQ-2, GAD-2, and PHQ-4 screening items ( $p < 0.01$ ).

## Discussion

This study aimed to assess the psychometric properties of PHQ-4 among out-of-school AGYW in Tanzania. We evaluated the reliability and validity of anxiety and depression screening scale (PHQ-4) among out of school AGYW. The findings from this study suggest that PHQ-4 can reliably and validly screen for core symptoms of depression and anxiety among AGYW. Results from CFA indicate that the PHQ-4 is a structurally valid, unidimensional measure of anxiety and depression with all items loaded

significantly into a single factor and loadings were ranging from 0.67 to 0.77. Principal component analysis with varimax-rotated component matrix ratified the original allocation of the items to the PHQ-4 scales, with the two anxiety items having the maximum factor loadings on component 1 and the two depression items having the maximum factor loadings on component 2. Furthermore, those who reported having suicidal thoughts and social function difficulties had statistically significantly higher scores on PHQ-2, GAD-2, and PHQ-4 screening items ( $p < 0.01$ ) than the others. Similar results were observed from a study conducted among primary-care patients in the United States and college students at a Midwestern university [9, 17].

The questionnaire was observed to be measuring a single construct, with all four questions having corrected item-total correlations above 0.3, which is the normally accepted cut-off level for keeping or discarding questionnaire items [38, 39]. Our findings were similar to a study on psychometric properties of PHQ-4 conducted among college students, although the correlation obtained was a bit higher compared to the current study [17]. In the current study, the internal consistency of PHQ-4 was supported by

**Table 3** Item Internal consistency of the scales

Over the last 2 weeks, how often have you been bothered by the following problems	<sup>a</sup> 4-Item alpha	2-Item alpha	<sup>b</sup> alpha if item- dropped
<b>Anxiety (GAD-2)</b>			
1. Feeling nervous, anxious or on edge	0.81	0.74	0.75
2. Not being able to stop or control worrying			0.74
<b>Depression (PHQ-2)</b>			
3. Little interest or pleasure in doing things		0.65	0.77
4. Feeling down, depressed or hopeless			0.77

<sup>a</sup>The overall alpha (Cronbach's  $\alpha$ : Indicates the internal reliability of the four items (conducted with the entire sample)); <sup>b</sup>Values of the overall alpha (Cronbach's  $\alpha$ ) if that item isn't included in the calculation

**Table 4** Principal Components Analysis and confirmatory factor analysis (CFA) for PHQ-4. (Factorial Validity)

Four-items of anxiety and depression scale	Component loadings		Goodness-of-fit statistics Confirmatory factor analysis
	<sup>a</sup> Component1	<sup>b</sup> Component2	
1. Feeling nervous, anxious or on edge	<b>0.834</b>	0.230	CFI = 0.995
2. Not being able to stop or control worrying	<b>0.799</b>	0.272	SRMR = 0.013
3. Little interest or pleasure in doing things	0.333	<b>0.716</b>	RMSEA = 0.054
4. Feeling down, depressed or hopeless	0.186	<b>0.865</b>	90%CI for RMSEA (0.031,0.079)

Note: Extraction method: principal components analysis. Rotation method: varimax;<sup>a</sup> Items on anxiety (GAD-2);

<sup>b</sup>Items on depression (PHQ-2); Components accounted for 82% of the variance

Cronbach's alpha ( $\alpha = 0.81$ ), which was above 0.70. However, for PHQ-2 and GAD-2 the Cronbach's alpha were a bit lower:  $\alpha = 0.65$ ,  $\alpha = 0.74$ , respectively. These findings were consistent but varied slightly from a study by Khubchandani et al. [17], who measured psychometric properties of PHQ-4 among college students ( $\alpha = 0.81$ ,  $\alpha = 0.76$ , and  $\alpha = 0.82$  for PHQ-4, PHQ-2 and GAD-2 respectively) and a study conducted by Lowe et al. among the general population in German ( $\alpha = 0.78$ ,  $\alpha = 0.75$ , and  $\alpha = 0.82$  for PHQ-4, PHQ-2 and GAD-2 respectively) [16]. The slight variations in the Cronbach's alpha between the current study and other previous validation studies conducted in Germany and USA [16, 17] may be caused by the difference in sample characteristics, setting and survey administration methods. However, all the values obtained by the current study are within acceptable reliability range [26].

The item response distributions were skewed to the lower scores, and floor effects were observed for all items. This means the present study sample response

**Table 5** Comparison of Mean Scores on Scales between Groups that have Depression and/ or Anxiety and reported suicidal thoughts and social function

	Two-item anxiety scale score (GAD-2)	Two-item depression scale score (PHQ-2)	Four-item scale score (PHQ-4)
	Mean( $\pm$ SE)	Mean( $\pm$ SE)	Mean( $\pm$ SE)
<sup>a</sup> Social function difficulties			
Yes	2.60 (0.04)	2.78 (0.04)	5.39 (0.07)
No	1.48 (0.06)	1.75 (0.05)	3.23 (0.09)
<sup>b</sup> Suicidal thoughts			
Yes	2.99 (0.06)	3.13 (0.06)	6.12 (0.09)
No	1.82 (0.04)	2.05 (0.04)	3.87 (0.07)

<sup>a</sup>How difficult have these problems (anxiety and/ or depression) made it for you to do your work, take care of things at home, or get along with other people? (No = Not difficult at all, yes = somewhat difficult /very difficult/ extremely difficult); Scores for individuals with social function problem were significantly higher compared to those who did not have ( $p < 0.01$ )

<sup>b</sup> Over the last 2 weeks, how often have you had thoughts that you would be better off dead or of hurting yourself in some way? (No = Not at all, yes = several days/more than half the days/nearly every day); Scores for individuals with suicidal thoughts were significantly higher compared to those who did not have ( $p < 0.01$ )

distribution is skewed towards no symptoms of anxiety or depression. The results of the floor effect could impair the impact of responsiveness and sensitivity [39, 40]. However, all of the alternative options in each item from the present study were selected, which point out that all response options were pertinent. Similar results were observed from a related study conducted using the Hospital Anxiety and Depression Scale (HADS) [41]. Besides, this study was based on data from the out-of-school AGYW population where a moderate proportion was shown to have symptoms of anxiety and depression. Thus, this problem might be related to the sample rather than the instrument.

#### Strengths and weaknesses

To our knowledge, this is the first study to assess the psychometric properties of the PHQ-4 scale in a large sample of out of school AGYW using ACASI. However, the scale has already been tested in other settings among samples with different characteristics [9, 17]. The current study sample was drawn from the community, unlike other studies where samples were drawn from healthcare facilities, clinics or the general adult population [9, 16].

This study has some limitations; the results are unlikely to be generalized to all AGYW in Tanzania since the sample was restricted to out of school AGYW from three districts in Shinyanga region in North West of Tanzania. Also, no clinical diagnostic evaluation was conducted as a reference standard. Thus, we cannot infer measures of screening accuracy, such as sensitivity and specificity for the PHQ-4. In addition, the screening diagnosis using PHQ-4 can have false-positive results since no strategies for confirming the diagnosis were used. This would indicate high levels of anxiety and depression without clinical problems [42, 43]. Furthermore, the current study drew data from a longitudinal study (CARE Study), in which the period between study data collection rounds was 6 months, therefore test-retest reliability could not be done. Nevertheless, the current study showed that the participants who screened positive for anxiety and depression were significantly more likely to report suicidal ideation and challenges with social interactions and functioning. Thus, the PHQ-4 can effectively be used for screening

core symptoms of depression and anxiety, and positive screens can be referred to the health facility for diagnosis and management as appropriate.

### Implications

AGYW who are out of school experience relatively more psychological distress compared to their school-attending peers [44]. As stated earlier, most mental health disorders are undiagnosed until adulthood [9–11]. As a result, the developmental potential, social life and economic impact of the AGYW are compromised [9–12]. This highlights the need for improving the screening and management of mental health disorders.

Utilizing a community sample, our assessment of PHQ-4 with out-of-school AGYW indicates that this tool can be used as a rapid, psychometrically sound and effective screening tool for depression and anxiety symptoms. The PHQ-4 is easy to administer, and the scoring system can easily be implemented through on-the-job training. Literature suggests that the decentralization of mental health services has the potential to expand coverage [8]. The tools such as the PHQ-4 can, therefore, be very useful in screening for anxiety and depression in the community, primary health facilities, research and programmatic settings [2, 9, 16]. This can be coupled with referral to higher-level facilities for diagnosis and management for those who screen positive.

### Conclusions

The findings from this study suggest that PHQ-4 item measure can reliably and validly measure core symptoms of depression and anxiety among the out of school AGYW. Overall, the PHQ-4 reliability and validity were reasonable and observed to be psychometrically sufficient and helpful for the measurement of anxiety and depression symptoms among AGYW. However, the problem with floor effects was observed for all items, although no serious impact on the factor structure observed due to this problem. But users need to be aware, as this problem could impair the impact of responsiveness and sensitivity. In spite of this, PHQ-4 can be endorsed to assess anxiety and depression symptoms among AGYW.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12888-020-02735-5>.

**Additional file 1.** Questionnaire

**Additional file 2.** Minimal dataset

### Abbreviations

ACASI: Audio computer assisted self-interviews; AGYW: Adolescent girls and young women; AUC: Area under the curve; CARE: Cash Transfer to Adolescent Girls and Young Women to reduce sexual risk behaviour;

CFA: Confirmatory factor analysis; CFI: Comparative Fit Index; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; GAD-2: Generalized Anxiety Disorders-2; GAD-7: Generalized Anxiety Disorders-7; GCP: Good clinical practice; HIV: Human Immunodeficiency Virus; IRB: Institutional Review Board; NIMR: National Institute for Medical Research; PCA: Principal components analysis; PHQ-2: Patient Health Questionnaire-2; PHQ-4: Patient Health Questionnaire-4; PHQ-9: Patient Health Questionnaire-9; PLHIV: People living with human immunodeficiency viruses; RMSEA: Root Mean Square Error of Approximation; SBCC: Social and behavior change communication; SD: Standard deviation; SRMR: Standardized Root Mean Square Residual

### Acknowledgments

The authors would like to thank the study participants, Sauti project management and Civil society organizations (CSO) working for the project in Kahama municipal council, Ushetu, and Msalala district councils. CARE study staff, especially Paul Moses, Peter Msofe and Eusebia Marandu of the National Institute for Medical Research, Mwanza Centre, and Esther Majani, Japhet Daud and Esabella Neeso of Jhpiego Tanzania- an affiliate of Johns Hopkins University, for their valuable input during the study.

### Ethical approval and consent to participate

CARE study was approved by the Medical Research Coordinating Committee of the National Institute for Medical Research (NIMR/HQ/R.8a/Vol.IX/2287) and the Johns Hopkins University Institutional Review Board (00007976). The trial is also registered at ClinicalTrials.gov, number NCT03597243. Each participant provided a written informed consent prior to study enrolment. Those who were illiterate provided a thumbprint. Participants who were at least 18 years provided voluntary informed consent for study participation. While those below 18 years assented and their parents or guardians provided informed consent for their participation in the study.

### Authors' contributions

JM, EK, DN, AG, AM, MK, AS, JC, SN and MW conceived and designed the study. JM, EK, DN and MW participated in data analysis and interpretation of the findings. JM, EK, DN, SN and MW drafted the manuscript while AG, AM, MK, AS and JC critically reviewed the manuscript for important intellectual content. All authors read and approved the final version of this manuscript.

### Funding

No funding was received for this particular study. This manuscript draws data from a cluster randomized controlled trial (CARE study) which is funded by a grant from the United States Agency for International Development (USAID) under Cooperative Agreement AID-621-A-15-00003. The funders had no role in the design, data collection and analysis, decision to publish, or preparation of this manuscript.

### Availability of data and materials

All data underlying the findings reported in this manuscript are provided as supporting information for this manuscript.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Sexual and Reproductive Health, National Institute for Medical Research, Mwanza Centre, P.O Box 1462, Mwanza, Tanzania. <sup>2</sup>Sauti Project, Jhpiego Tanzania - an affiliate of Johns Hopkins University, Dar es Salaam, Tanzania.

Received: 19 October 2019 Accepted: 15 June 2020

Published online: 19 June 2020

### References

1. Organization WH: Global Health Estimates 2015: disease burden by cause, age, sex, by country and by region, 2000–2015. Geneva, 2016; 2018.
2. Mathur S, Okal J, Musheke M, Pilgrim N, Patel SK, Bhattacharya R, Jani N, Matheka J, Banda L, Mulenga D. High rates of sexual violence by both

- intimate and non-intimate partners experienced by adolescent girls and young women in Kenya and Zambia: findings around violence and other negative health outcomes. *PLoS One*. 2018;13(9):e0203929.
3. Lwidiko A, Kibusi SM, Nyundo A, Mpondo BC. Association between HIV status and depressive symptoms among children and adolescents in the southern highlands zone, Tanzania: a case-control study. *PLoS One*. 2018;13(2):e0193145.
  4. Himle JA, Baser RE, Taylor RJ, Campbell RD, Jackson JS. Anxiety disorders among African Americans, blacks of Caribbean descent, and non-Hispanic whites in the United States. *J Anxiety Dis*. 2009;23(5):578–90.
  5. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. *Lancet*. 2007;369(9569):1302–13.
  6. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustun TB. Age of onset of mental disorders: a review of recent literature. *Curr Opin Psychiatry*. 2007;20(4):359.
  7. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. *Lancet*. 2007;370(9590):878–89.
  8. Saraceno B, van Ommeren M, Batniji R, Cohen A, Gureje O, Mahoney J, Sridhar D, Underhill C. Barriers to improvement of mental health services in low-income and middle-income countries. *Lancet*. 2007;370(9593):1164–74.
  9. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*. 2009;50(6):613–21.
  10. Hendriks SM, Spijker J, Licht CM, Hardevelde F, de Graaf R, Batelaan NM, Penninx BW, Beekman AT. Long-term work disability and absenteeism in anxiety and depressive disorders. *J Affect Disord*. 2015;178:121–30.
  11. de Lijster JM, Dieleman GC, Utens EM, Dierckx B, Wierenga M, Verhulst FC, Legerstee JS. Social and academic functioning in adolescents with anxiety disorders: a systematic review. *J Affect Disord*. 2018;230:108–17.
  12. Wang X, Liu Z, Li Y, Li G, Huang Y. Association of comorbidity of mood and anxiety disorders with suicidal behaviors. *J Affect Disord*. 2018;227:810–6.
  13. Liu Y, Kaaya S, Chai J, McCoy D, Surkan P, Black M, Sutter-Dallay A-L, Verdoux H, Smith-Fawzi M. Maternal depressive symptoms and early childhood cognitive development: a meta-analysis. *Psychol Med*. 2017;47(4):680–9.
  14. Kakuma R, Minas H, van Ginneken N, Dal Poz MR, Desiraju K, Morris JE, Saxena S, Scheffler RM. Human resources for mental health care: current situation and strategies for action. *Lancet*. 2011;378(9803):1654–63.
  15. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res*. 2005;58(2):163–71.
  16. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, Schneider A, Brähler E. A 4-item measure of depression and anxiety: validation and standardization of the patient health Questionnaire-4 (PHQ-4) in the general population. *J Affect Disord*. 2010;122(1–2):86–95.
  17. Khubchandani J, Brey R, Kotecki J, Kleinfelder J, Anderson J. The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Arch Psychiatr Nurs*. 2016;30(4):457–62.
  18. Kuringe E, Materu J, Nyato D, Majani E, Ngeni F, Shao A, Mjungu D, Mtenga B, Nnko S, Kipingili T. Prevalence and correlates of depression and anxiety symptoms among out-of-school adolescent girls and young women in Tanzania: a cross-sectional study. *PLoS One*. 2019;14(8):e0221053.
  19. Estes LJ, Lloyd LE, Teti M, Raja S, Bowleg L, Allgood KL, Glick N. Perceptions of audio computer-assisted self-interviewing (ACASI) among women in an HIV-positive prevention program. *PLoS One*. 2010;5(2).
  20. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–13.
  21. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293–307.
  22. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
  23. Ritter NL. Understanding a widely misunderstood statistic: Cronbach's. *Online Submission*; 2010.
  24. Fayers P, Machin D. *Quality of Life Assessment, Analysis and Interpretation*. Chichester, England: Wiley; 2000.
  25. Bland JM, Altman DG. *Statistics notes: Cronbach's alpha*. *Bmj*. 1997;314(7080):572.
  26. Nunnally JC. *Psychometric theory*. 2d Ed: McGraw-hill; 1978.
  27. Lance CE, Butts MM, Michels LC. The sources of four commonly reported cutoff criteria: what did they really say? *Organ Res Methods*. 2006;9(2):202–20.
  28. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. *Practical Assessment Res Eval*. 2012;17:3.
  29. Lee SY, Poon WY, Bentler PM. A two-stage estimation of structural equation models with continuous and polytomous variables. *Br J Math Stat Psychol*. 1995;48(2):339–58.
  30. O'Rourke N, Hatcher L. *A step-by-step approach to using SAS for factor analysis and structural equation modeling*: SAS institute; 2013.
  31. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J*. 1999;6(1):1–55.
  32. Browne MW, Cudeck R. *Alternative ways of assessing model fit*. Sage Focus Editions. 1993;154:136.
  33. Byrne BM. *Structural equation modeling with AMOS: basic concepts, applications, and programming*: Routledge; 2016.
  34. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods*. 1996;1(2):130.
  35. Li J, Linear RR. *Principal component analysis*; 2014.
  36. Jolliffe I. *Principal component analysis*: springer; 2011.
  37. Price JH, Braun RE, Khubchandani J, Payton E, Bhattacharjee P. Development of an attribution of racial/ethnic health disparities scale. *J Community Health*. 2014;39(4):792–9.
  38. De Vaus D. *Analyzing social science data: 50 key problems in data analysis*: sage; 2002.
  39. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. USA: Oxford University Press; 2015.
  40. Bennett SJ, Oldridge NB, Eckert GJ, Embree JL, Browning S, Hou N, Deer M, Murray MD. Discriminant properties of commonly used quality of life measures in heart failure. *Qual Life Res*. 2002;11(4):349–59.
  41. Djukanovic I, Carlsson J, Årestedt K. Is the hospital anxiety and depression scale (HADS) a valid measure in a general population 65–80 years old? A psychometric evaluation study. *Health Qual Life Outcomes*. 2017;15(1):193.
  42. Klein MC, Ciotoli C, Chung H. Primary care screening of depression and treatment engagement in a university health center: a retrospective analysis. *J Am Coll Heal*. 2011;59(4):289–95.
  43. Shean G, Baldwin G. Sensitivity and specificity of depression questionnaires in a college-age sample. *J Genet Psychol*. 2008;169(3):281–92.
  44. Rock A, Barrington C, Abdoulayi S, Tsoka M, Mvula P, Handa S. Social networks, social participation, and health among youth living in extreme poverty in rural Malawi. *Soc Sci Med*. 2016;170:55–62.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

