RESEARCH

Open Access

A Chinese adaptation of the Patient Health Questionnaire for Adolescents (PHQ-A): factor structure and psychometric properties

Yu-Qi Chen[1,2], Xiao-Jie Huang[2], Fan Yang[2], Jing-Jing Yang[2], Jing Zhong[1,2], Kai-Min Yao[1,2], Jing-Xiao Kuang[1,2] and Ming-Zhi Xu[1,2]*

Abstract

Background To examine the factor structure and psychometric properties of the Patient Health Questionnaire for Adolescents (PHQ-A) in Chinese children and adolescents with major depressive disorder (MDD).

Methods A total of 248 MDD patients aged between 12 and 18 years were recruited and evaluated by the Patient Health Questionnaire for Adolescents (PHQ-A), the Center for Epidemiological Survey Depression Scale (CES-D), the Mood and Feelings Questionnaire (MFQ), and the improved Clinical Global Impression Scale, Severity item (iCGI-S). Thirty-one patients were selected randomly to complete the PHQ-A again one week later. Confirmatory factor analysis (CFA) was used to test the construct validity of the scale. Reliability was evaluated by Macdonald Omega coefficient. Pearson correlation coefficient was used to assess the item-total correlation and the correlation of PHQ-A with CES-D and MFQ respectively. Spearman correlation coefficient was used to assess test-retest reliability. The optimal cut-off value, sensitivity, and specificity of the PHQ-A were achieved by estimating the Receiver Operating Characteristics (ROC) curve.

Results CFA reported adequate loadings for all items, except for item 3. Macdonald Omega coefficient of the PHQ-A was 0.87. The Spearman correlation coefficient of the test-retest reliability was 0.70. The Pearson correlation coefficients of the PHQ-A with CES-D and MFQ were 0.87 and 0.85, respectively ($p < 0.01$). By taking the iCGI-S as the remission criteria for MDD, the optimal cut-off value, sensitivity and specificity of the PHQ-A were 7, 98.7%, 94.7% respectively.

Conclusion The PHQ-A presented as a unidimensional construct and demonstrated satisfactory reliability and validity among the Chinese children and adolescents with MDD. A cut-off value of 7 was suggested for remission.

Keywords Depression, PHQ-A, Adolescents, Psychometric properties, Factor structure, Cut-off value

*Correspondence:
Ming-Zhi Xu
xumingzhi@gdph.org.cn
[1] School of Public Health, Southern Medical University, Guangzhou, Guangdong, People's Republic of China
[2] Guangdong Mental Health Center, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong 510120, People's Republic of China

Introduction

Major Depression Disorder (MDD) is a prevalent mental disorder. Studies in various countries have reported its lifetime prevalence of, for example, 20.5% in Chile [1], 21% in France [2], 6.7% in South Korea [3] and 3.4% in China [4]. In 2008, the World Health Organization (WHO) ranked MDD as the third leading contributor to the global burden of disease, predicting that it would ascend to the primary position by 2030 [5, 6]. Among

Chen *et al. BMC Psychiatry*   (2024) 24:331

Page 2 of 8

Chinese children and adolescents, depressive symptoms have become more prevalent and the pooled prevalence of depressive symptoms among them was 22.2% [7]. Depression often leads to a cascade of consequences, including suicidal ideation, school dropout, behavioral disturbances, and substance abuse among children and adolescents [8]. Additionally, 10%-25% of patients with mild symptoms are at risk of scoring in the severe range over one to three years without intervention in time [9]. Adolescence is frequently regarded as a pivotal period for the early identification and prevention of adult depression [10], highlighting the vital importance of prompting diagnosis and intervention to prevent the onset of depression in adolescents.

Using simple and efficacious screening tools is an effective way to improve the detection of mental disorders [11]. The US Preventive Services Task Force (USPSTF) recommended to adopt screening tools like the Patient Health Questionnaire for Adolescents (PHQ-A) and the primary care version of the Beck Depression Inventory (BDI) for the identification of depression among adolescents in primary care settings [12]. Self-rating scales are simple and efficacious screening tools to screen depression among adolescents. It can not only help clinicians to quantify the patients' subjective feelings, but also enable patients understand the severity of their distress. Meanwhile, the Measurement-based care (MBC), which focus on the periodical assessments of the treatment in the process of quality controls, has been widely recommended in psychiatric practice. The MBC allows clinicians to make personalize treatment decisions for the patients, thereby improving the adoption of appropriate treatment strategies, reducing treatment resistance, and increasing treatment quality [13].

In clinical practice, the depression assessment scales used for children and adolescents mainly include the Children's Depression Inventory (CDI), the Center for Epidemiological Studies Depression Scale (CES-D), the primary care version of the Beck Depression Inventory (BDI), the Mood and Feelings Questionnaire (MFQ) and the Patient Health Questionnaire for Adolescents (PHQ-A/PHQ-9 M) [12, 14].

The PHQ-A/PHQ-9 M is adapted from the 9-item Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 was initially developed by Kroenke and Spitzer in 2001 for assessing depression in adult primary care and then was extended to adolescent depression [15–18]. The advantage of the PHQ-9 was that it exclusively focuses on the 9 diagnostic criteria of MDD in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), which makes it more specific for major depression and may accurately discriminate depression from anxiety or even general psychological

distress [19]. However, previous studies reported that the specificity of the PHQ-9 was lower when used in adolescents, which could lead to an increased likelihood of false positive rate [16, 20], suggesting that the PHQ-9 may not be the most suitable screening tool for assessing depression in adolescents. This could potentially be attributed to the specific features of adolescent depression that may differ from adult depression. Although some of the symptoms of MDD may be similar for adults and adolescents, its clinical feature and prominence may be different [21]. Irritability is an impairing clinic manifestation that refers to easy annoyance and touchiness, which has been the most frequently reported symptom in adolescent depression [21, 22]. Although the DSM-IV identifies irritability as a characteristic of adolescent depression, it states that 'it is not a criterion for major depressive episode' [21]. Yet both the DSM-5 and the 11th edition of the International Classification of Diseases (ICD-11) include the criteria to specifically define MDD for adolescent with the statement 'Depressed mood (subjective or observed) can be irritable mood in children and adolescents' [23], suggesting the importance of irritability on the diagnosis of MDD among adolescents. Comparing to the PHQ-9, the PHQ-A adds 'irritability' in the description of item 1, which is the major revision. Moreover, two minor revisions include that the order of item4 (Fatigability) and item5 (Appetitive problems) reverse, and the PHQ-A added 'school work' in item7 (Concentration problems), which adapts to children's daily activities. With these revisions, the PHQ-A may be more appropriate for screening depression in adolescents and adjusts the DSM-5 and ICD-11 diagnosis better.

The English version of the PHQ-A showed satisfactory sensitivity, specificity and overall diagnostic accuracy and its reliability and validity was proved satisfactory [24, 25]. The PHQ-A has been translated into other languages such as Urdu, Thai, Portuguese and Arabic [26–29]. All of these translations demonstrated satisfactory reliability and validity in relative countries and populations. In the Urdu, Thai, and Arabic version, only one factor was extracted by exploratory factor analysis (EFA). Furthermore, the unidimensional factor structure is verified by confirmatory factor analysis (CFA) in the Urdu and Arabic version [26–29]. Therefore, the PHQ-A is a promising screening tool and merits further evaluation among adolescents in China.

In this study, we hypothesized that the PHQ-A would fit into a unidimensional structure and demonstrate good psychometric properties among Chinese children and adolescents with MDD. We would test the hypotheses and recommend a cut-off value for remission.

Chen *et al. BMC Psychiatry*   (2024) 24:331

Page 3 of 8

Method

Participants

Participants were recruited from the outpatient and inpatient departments of Guangdong mental health center, Guangdong Provincial People's Hospital from December 2022 to June 2023. This study was authorized by the Clinical Research Ethics Committee of Guangdong Provincial People's Hospital (No. KY-Z-2022-062-01), and carried out based on the Declaration of Helsinki. Informed consent was obtained in all patients. The enrollment criteria were as follows: (1) sign an informed consent form to participate in the research; (2) be between 12 and 18 years old; (3) meet the diagnostic criteria of MDD in the DSM-5; (4) be able to understand the questionnaire and report their status. The exclusion criteria were: (1) diagnosed with bipolar disorder or other severe mental disorders such as schizophrenia; (2) a history of alcoholism, drug abuse, or serious physical diseases; (3) diagnosed with mental retardation. Recommendations in the literature suggested a minimum sample size for CFA of around 100-250 observations [30].

Patients who met the conditions for enrollment entered the study group. Their demographic data (age, birth date, sex, ethnicity, education year, living conditions, insurance type, age of first onset, and family history of mental illness) were collected. Then they completed the scales in a quiet room. 31 patients of the participants were randomly selected and retested 7 days after the first test to assess the test-retest reliability of the PHQ-A.

Instruments

The Patient Health Questionnaire for Adolescents (PHQ-A)

The PHQ-A is adapted from the PHQ-9 for measuring depression among adolescents. It is a 9-item self-report scale developed by Spitzer and Johnson in 2002. The items are rated on a 4-point Likert scale: 0 (none), 1 (several days), 2 (more than half the days) and 3 (nearly every day). The total score ranges from 0 to 27. Higher scores indicate greater symptom severity. The PHQ-A is an acceptable and efficient screening tool in primary care settings [24].

In this study we translated the PHQ-A and the translating process were as follows. Firstly, one psychiatrist and two psychologists who were good at English translated the original version of PHQ-A from English to Chinese. Secondly, after translating it back into English, they made adjustment until it was almost the same as the original one. Thirdly, we choose 10 individuals at random to participate in a pilot test. Based on the feedback from the test, the final Chinese translation was confirmed.

The Center for Epidemiological Survey Depression Scale (CES-D)

The CES-D was used as the criterion scale in our study. It is a 20-item self-rating scale developed by Radloff in 1977 [31]. The CES-D are rated on a 4-point Likert scale. The item 4, the item 8, the item 12 and the item 16 are reversely scored. The total score ranges from 0 to 60, with higher scores suggesting greater symptom severity. The Chinese version of CES-D is a reliable and valid screening tool for Chinese adolescents [32–34].

The Mood and Feelings Questionnaire (MFQ)

The Mood and Feelings Questionnaire (MFQ) was used as another criterion scale in our study. The child self-report MFQ developed by Angold in 1995 is a 33-item scale [35]. It was developed to measure depression among children and adolescents. The child self-report MFQ is rated on a 3-point Likert scale: 0 (not true) to 2 (true), with higher scores suggesting greater depressive symptom severity. The child self-report MFQ is a reliable and valid screening tool for Chinese children and adolescents [36].

The improved Clinical Global Impression Scale, Severity item (iCGI-S)

The Clinical Global Impression (CGI) scale is a classic scale to rate the clinician's overall impression upon a patient's current condition [37]. The iCGI-S is an item of improved CGI that rates the severity of depression symptoms. It is rated by clinicians on a 7-point Likert scale: 1=normal, 2=borderline ill, 3=mildly ill, 4=moderately ill, 5=markedly ill, 6=severely ill, and 7=among the most extremely ill. iCGI-S=1 was commonly used to differentiate patients in remission from those are not and acted on establishing the cut-off score [38].

Statistical analysis

EpiData3.1 was used for data entry, AMOS 24.0 for verifying the factor structure of the PHQ-A and SPSS 25.0 for data processing and analysis. Confirmatory factor analysis (CFA) was used to verify the unidimensional structure of the PHQ-A. Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Goodness of Fit Index (GFI) and Root Mean Square Error of Approximation (RMSEA) were selected as indicators. Based on the existing recommendations, acceptable model fit was considered if the CFI, GFI and TLI were greater than 0.9 and the RMSEA was smaller than 0.10 [39–42]. Macdonald Omega coefficient was used to evaluate the internal consistency reliability. Spearman correlation coefficient was applied to calculate the test-retest reliability. Item-total correlation and criteria validity were estimated via Pearson correlation

Chen *et al. BMC Psychiatry*   (2024) 24:331

Page 4 of 8

coefficient. Due to the sample size, Kruskal–Wallis H Test was used to compare the total scores of the PHQ-A with different severity of depressive symptoms classified by the iCGI-S. Receiver Operating Characteristic (ROC) curve was established on the basis of iCGI-S = 1. Area Under the Curve (AUC) and Youden's index were calculated to evaluate the sensitivity, specificity and optimal cut-off score of the PHQ-A.

Results

Demographic characteristics and distribution of the PHQ-A scores

A total of 253 patients were recruited. Five patients who did not perform seriously (such as responding in a certain order of categories like '1,2,3,4,1,2,3,4……' or responding rapidly and finishing the scales in an unreasonable time) were considered to be invalid responses and were eliminated. Finally, 248 patients were included with a response

Table 1 Demographic characteristics of the participants

	Total Sample (*n* = 248)	%
Age		
12–15	125	50.4
16–18	123	49.6
Gender		
Female	187	75.4
Male	61	24.6
First episode		
Yes	212	85.5
No	36	14.5
Ethnicity		
Han	242	97.6
Others	6	2.4
Living condition		
Alone	4	1.6
With family	231	93.2
With others	13	5.2
Insurance type		
Yes	76	30.6
No	172	69.4
Family history of mental disorders		
Yes	36	14.5
No	212	85.5
Edu year		
5–9	131	52.8
10–12	117	47.2
ICGI-S score		
1	19	7.7
≥2	229	92.3
Mean score	16.81	SD = 6.23

Note: *Mean score* the mean score of the PHQ-A, *SD* Standard Deviation

rate of 98.02%. As shown in Table 1, the age of the patients ranged from 12 to 18 years, with a mean age of 15.42 years (SD = 1.70). 187 patients (75.4%) were female and the other 61 (24.6%) were male. The age on their first episode ranged from 8 to 18 years, with an average of 14.60 (SD = 1.82). 212 patients (85.5%) experienced their first MDD episode. Their education years ranged from 5 to 12 years, with an average of 9.35 (SD = 1.75). The mean score of PHQ-A in females and males was 16.98 (SD = 6.16) and 16.28 (SD = 6.43) respectively, which did not differ significantly (t = 0.77 and *P* = 0.44).

The mean score of the PHQ-A of the total sample was 16.81 (SD = 6.23). The score of the patients ranged from 0 to 27 with a mode of 20. According to the data, only 8 participants (3.2%) received a minimum score, while 42 (9.1%) had the maximum score of 27.

Construct validity

AMOS24.0 was used to conduct the confirmatory factor analysis (CFA), which verified the unidimensional structure of the PHQ-A. As demonstrated in Fig. 1, except for the item3 (lamma = 0.47), all items showed relatively adequate loadings on the latent factor (t > 1.96). The goodness of fit indices of the model was within
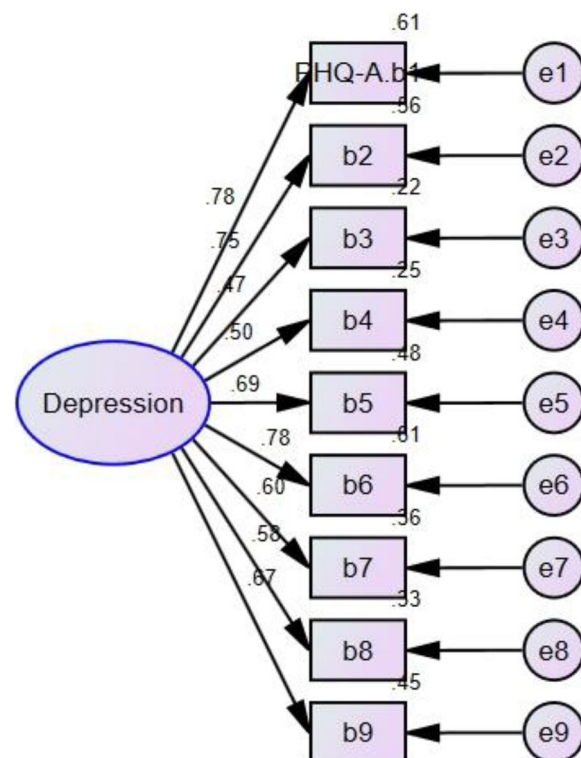

Fig. 1 Single-factor CFA analysis to the Chinese translation of the Patient Health Questionnaire for Adolescents (PHQ-A). (Note: b1 = item1 of the PHQ-A; b2 = item2; b3 = item3; b4 = item4; b5 = item5; b6 = item6; b7 = item7; b8 = item8; b9 = item9)

Chen *et al. BMC Psychiatry*   (2024) 24:331

Page 5 of 8

the ideal parameter range. The statistical parameters were as follows: Tucker-Lewis index (TLI) = 0.92, Comparative Fit Index (CFI) = 0.94, Goodness of Fit Index (GFI) = 0.94, Root Mean Square Error of Approximation (RMSEA) = 0.08.

Reliability analysis

Macdonald Omega coefficient of the PHQ-A was 0.87, demonstrating acceptable internal consistency of the PHQ-A.

The PHQ-A has acceptable test-retest reliability with a Spearman correlation coefficient of 0.70 ($p < 0.01$), showing acceptable stability across time within the scale [43].

Item-total correlation analysis of the PHQ-A

Pearson correlation analysis was used to calculate the item-total correlation coefficients of the PHQ-A. As demonstrated in Table 2, the item-total correlation coefficients of the scale ranged from 0.45 to 0.71 ($p < 0.01$). The numerical value of the correlation coefficients between the items varied ranged from 0.29 (item 4 and item 5) to 0.68 (item 1 and item 6), showing that item-total score correlations were all significant ($p < 0.01$).

Concurrent validity

The Pearson correlation coefficient of the PHQ-A with the CES-D and MFQ was 0.87 and 0.85 ($p < 0.01$), respectively. The result demonstrated that the PHQ-A has good concurrent validity with these scales.

The patients were classified into different groups by the score of the iCGI-S. Their scores ranged from 1 to 6 (No patient was rated as 7). As shown in Table 3, the Kruskal-Wallis H Test revealed that on the total scores of the PHQ-A, Group 1 was significantly smaller than Group 4, Group 5 and Group 6; Group 3 was significantly smaller than Group 4, Group 5 and Group 6; Group 4

Table 3 Kruskal–Wallis H Test among subgroups with different iCGI-S scores

Group comparison	Std.TS	SD	*p* value
group1-group2	-0.519	39.396	1.000
group1-group3	-1.087	22.201	1.000
group1-group4	-4.713	18.497	**0.000**
group1-group5	-8.898	17.653	**0.000**
group1-group6	-6.316	30.182	**0.000**
group2-group3	-0.095	38.785	1.000
group2-group4	-1.813	36.387	1.000
group2-group5	-3.755	36.384	**0.003**
group2-group6	-3.881	43.854	**0.002**
group3-group4	-3.670	17.182	**0.004**
group3-group5	-8.172	16.269	**0.000**
group3-group6	-5.665	29.394	**0.000**
group4-group5	-6.549	10.674	**0.000**
group4-group6	-3.874	26.707	**0.002**
group5-group6	-1.285	26.129	1.000

Note: *Std.TS* Standard test statistic, *SD* Standard Deviation

Values in boldface indicate statistical significance

was significantly smaller than Group 5 and Group 6. The results demonstrated that the PHQ-A was able to successfully discriminate patients with different severity of depressive symptoms.
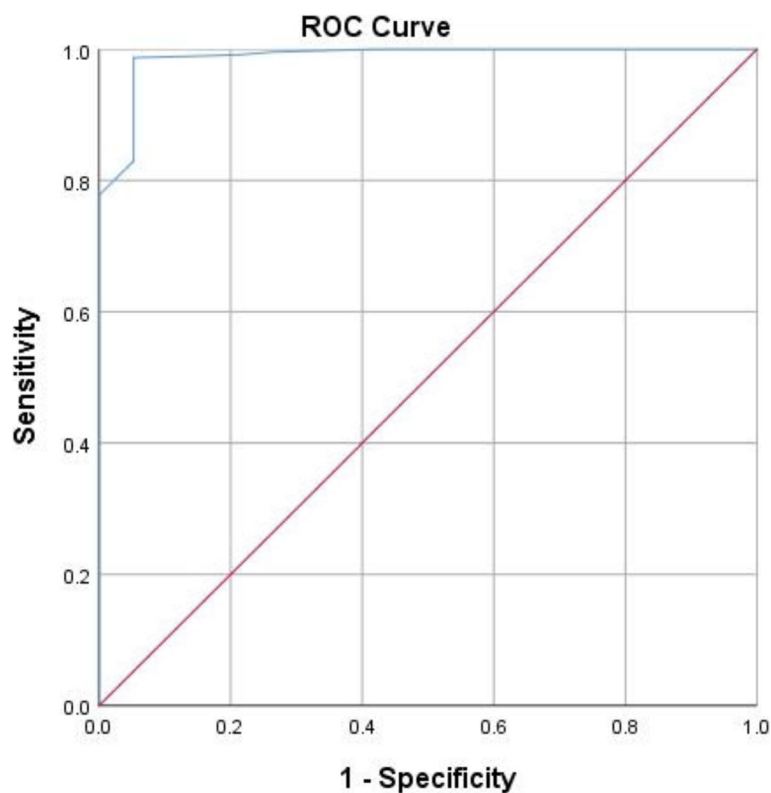
Cut-off value

Remission of symptoms is the ideal result of treatment for mood disorders, signifying patient recovery from mood disorders [44, 45]. We took iCGI-S = 1 as the criteria for MDD remission and SPSS 25.0 was used to draw the ROC curve of the PHQ-A. As shown in Fig. 2, the areas under the ROC curve (AUC) of the scale were 0.99 (95% CI: 0.97–1.00, $P < 0.01$). The maximum Youden's index was 0.93, identifying the best cutoff value of 7.5. While the cutoff value was used as an integer, it should be recommended to be 7 or 8 for the PHQ-A. 7 could be a stricter criterion for remission which was clinically helpful. At this point with a sensitivity of 98.7% and specificity of 94.7%, the scale had the best distinction ability.

Discussion

We investigated the psychometric properties of the PHQ-A among Chinese children and adolescents with MDD in the study.

The CFA showed that except for the item 3 (Sleeping problems), all items showed relatively adequate loadings on the latent factor. This finding of the relatively lower loading on item 3 was consistent with the study of the Jordan version of the PHQ-A [26], suggesting that the item 3 might be less predictive. Generally, all

Table 2 Item-total correlation and factor loading for PHQ-A

	Item-total correlation	CFA Factor Loading
Item1 Lack of interest	0.71	0.78
Item2 Depressed mood	0.67	0.75
Item3 Sleeping problems	0.45	0.47
Item4 Appetitive problems	0.49	0.50
Item5 Fatigability	0.64	0.69
Item6 Negative feelings about self	0.69	0.78
Item7 Concentration problems	0.56	0.60
Item8 Psychomotor agitation / retardation	0.56	0.58
Item9 Suicidal ideation	0.61	0.67

Note: *CFA* confirmatory factor analysis

Chen *et al. BMC Psychiatry*   (2024) 24:331

Page 6 of 8

Fig. 2 The receiver operating characteristic (ROC) curve of the Chinese translation of the Patient Health Questionnaire for Adolescents (PHQ-A)

the goodness-of-fit indices of the model fell within the acceptable parameter range, indicating acceptable structural validity of the scale. The CFA of the PHQ-A confirmed its unidimensional structure, which is consistent with previous research findings [26–29].

In terms of the reliability of the PHQ-A, the internal reliability was examined by Macdonald Omega coefficient. The value of the coefficient was 0.87, indicating good internal consistency and high item homogeneity within the PHQ-A. These findings are consistent with previous research on the Arabic, Urdu, Mozambique, and Thai versions of the PHQ-A [26–29]. The Spearman correlation coefficient of the scale was 0.70, which suggested good test–retest reliability for the scale, indicating the stability of the PHQ-A. These findings align with prior studies [27–29]. The satisfactory internal reliability and fair test-retest reliability of the PHQ-A confirmed its acceptable reliability among Chinese adolescents with MDD.

For criterion validity, the CES-D and MFQ were selected as the criterion scales. The total score of the PHQ-A was highly positively correlated with CES-D and MFQ, indicating that the PHQ-A had good criterion validity with them. The results of the Kruskal-Wallis H Test demonstrated that the PHQ-A could successfully discriminate different levels of severity of depressive symptoms among Chinese children and adolescents, consistent with the results of earlier research [25, 26, 28, 29].

This study showed that the area under the ROC curve (AUC) was 0.99, suggesting that the PHQ-A might be a valid tool. However, the AUC of the Thai and Mozambique versions were 0.88 and 0.85 respectively [27, 29]. The very high AUC value in this study might be due to the small sample size of remission ($N = 19$). The optimal cut-off score for the PHQ-A was 7 (7 will be a stricter criterion for remission which is clinically helpful), conforming with the recommended scores from previous studies on other PHQ-A versions [25, 26, 28, 29]. Both the Thai and Mozambique versions of the PHQ-A had an optimal cutoff value of 8. The sensitivity and specificity of the Thai version of PHQ-A were 76% and 81% [29]. The sensitivity and specificity of the Portuguese version were 78% and 80% [27]. Both were lower than the results in our study.

The results indicated that the psychometric properties, the factor structure and the optimal cut-off score of the PHQ-A performed relatively stable across different cultures and languages. There are several limitations in this study. Firstly, since our study focused on only

Chen *et al. BMC Psychiatry*   (2024) 24:331

Page 7 of 8

MDD patients, its findings might not apply to other clinical population. Secondly, the sample was recruited from a single mental health center, which was less ideal than multi-center recruitment. Thirdly, this research is cross-sectional, which is unable to confirm whether the scale is sensitive to changes of symptoms after treatment. Finally, the relatively small sample size of remission might cause bias in AUC. Therefore, future studies are needed to test the psychometric characteristics of the PHQ-A in different populations and verify its sensitivity to treatment changes.

Conclusion

The PHQ-A has satisfactory psychometric properties on Chinese children and adolescents with MDD. It is a simple, reliable, and valid measuring tool for screening and assessing the severity of MDD symptoms among children and adolescents. In this study, it is recommended that the optimum cut-off value of the scale is 7.

Authors' contributions
Yu-Qi Chen: Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. Xiao-Jie Huang: Investigation, Data curation, Validation, Writing – review & editing. Fan Yang: Investigation, Funding acquisition. Jing-Jing Yang: Resources, Validation. Jing Zhong: Validation. Kai-Min Yao: Writing – review & editing. Jing-Xiao Kuang: Writing – review & editing. Ming-Zhi Xu: Conceptualization, Validation, Methodology, Resources, Writing – review & editing, Supervision, Project administration. All authors reviewed the manuscript.

Availability of data and materials
The datasets generated and analyzed during the current study are not publicly available due to limitations written into the participants' consent forms. Group level data are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
This study protocol was approved by the Clinical Research Ethics Committee of Guangdong Provincial People's Hospital (No. KY-Z-2022–062-01). This study was performed in accordance with the principles expressed in the Declaration of Helsinki. All participants have provided written informed consent to participate in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Markkula N, Zitko P, Peña S, Margozzini P, Retamal CP. Prevalence, trends, correlates and treatment of depression in Chile in 2003 to 2010. Soc Psychiatry Psychiatr Epidemiol. 2017;52(4):399–409.
2. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, de Graaf R, Demyttenaere K, Hu C, Iwata N, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. 2011;9:90.
3. Cho MJ, Seong SJ, Park JE, Chung IW, Lee YM, Bae A, Ahn JH, Lee DW, Bae JN, Cho SJ, et al. Prevalence and Correlates of DSM-IV Mental Disorders in South Korean Adults: The Korean Epidemiologic Catchment Area Study 2011. Psychiatry Investig. 2015;12(2):164–70.
4. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, Yu Y, Kou C, Xu X, Lu J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry. 2019;6(3):211–24.
5. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry. 2022;9(2):137–50.
6. Malhi GS, Mann JJ. Depression. Lancet. 2018;392(10161):2299–312.
7. Li JY, Li J, Liang JH, Qian S, Jia RX, Wang YQ, Xu Y. Depressive Symptoms Among Children and Adolescents in China: A Systematic Review and Meta-Analysis. Med Sci Monit. 2019;25:7459–70.
8. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry. 2015;56(3):345–65.
9. Cuijpers P, Smit F. Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. Acta Psychiatr Scand. 2004;109(5):325–31.
10. Kleinert S. Adolescent health: an opportunity not to be missed. Lancet. 2007;369(9567):1057–8.
11. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care. 2004;42(12):1194–201.
12. Siu AL. Screening for Depression in Children and Adolescents: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016;164(5):360–6.
13. Adli M, Bauer M, Rush AJ. Algorithms and collaborative-care systems for depression: are they effective and why? A systematic review. Biol Psychiatry. 2006;59(11):1029–38.
14. Forman-Hoffman VL, Viswanathan M. Screening for Depression in Pediatric Primary Care. Curr Psychiatry Rep. 2018;20(8):62.
15. Allgaier AK, Pietsch K, Frühe B, Sigl-Glöckner J, Schulte-Körne G. Screening for depression in adolescents: validity of the patient health questionnaire in pediatric care. Depress Anxiety. 2012;29(10):906–13.
16. Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, Rockhill C, Katon W. Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents. Pediatrics. 2010;126(6):1117–23.
17. Fonseca-Pedrero E, Díez-Gómez A, Pérez-Albéniz A, Al-Halabí S, Lucas-Molina B, Debbané M. Youth screening depression: Validation of the Patient Health Questionnaire-9 (PHQ-9) in a representative sample of adolescents. Psychiatry Res. 2023;328: 115486.
18. Kroenke K. PHQ-9: global uptake of a depression scale. World psychiatry. 2021;20(1):135–6.
19. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. Psychiatr Ann. 2002;32(9):509–15.
20. Ganguly S, Samanta M, Roy P, Chatterjee S, Kaplan DW, Basu B. Patient health questionnaire-9 as an effective tool for screening of depression among Indian adolescents. J Adolesc Health. 2013;52(5):546–51.
21. Crowe M, Ward N, Dunnachie B, Roberts M. Characteristics of adolescent depression. Int J Ment Health Nurs. 2006;15(1):10–8.
22. Stringaris A. Irritability in children and adolescents: a challenge for DSM-5. Eur Child Adolesc Psychiatry. 2011;20(2):61–6.

Chen *et al. BMC Psychiatry*   (2024) 24:331

Page 8 of 8

23. Uher R, Payne JL, Pavlova B, Perlis RH. Major depressive disorder in DSM-5: implications for clinical practice and research of changes from DSM-IV. Depress Anxiety. 2014;31(6):459–71.
24. Johnson JG, Harris ES, Spitzer RL, Williams JB. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolescent Health. 2002;30(3):196–204.
25. Nandakumar AL, Vande Voort JL, Nakonezny PA, Orth SS, Romanowicz M, Sonmez AI, Ward JA, Rackley SJ, Huxsahl JE, Croarkin PE. Psychometric Properties of the Patient Health Questionnaire-9 Modified for Major Depressive Disorder in Adolescents. J Child Adolesc Psychopharmacol. 2019;29(1):34–40.
26. Al-Amer R, Maneze D, Ramjan L, Villarosa AR, Darwish R, Salamonson Y. Psychometric testing of the Arabic version of the Patient Health Questionnaire among adolescent refugees living in Jordan. Int J Ment Health Nurs. 2020;29(4):685–92.
27. Lovero KL, Adam SE, Bila CE, Canda ED, Fernandes ME, Rodrigues TIB, Sander MCT, Mellins CA, Duarte CS, Dos Santos PF, et al. Validation of brief screening instruments for internalizing and externalizing disorders in Mozambican adolescents. BMC Psychiatry. 2022;22(1):549.
28. Naveed S, Waqas A, Memon AR, Jabeen M, Sheikh MH. Cross-cultural validation of the Urdu translation of the Patient Health Questionnaire for Adolescents among children and adolescents at a Pakistani school. Public Health. 2019;168:59–66.
29. Panyawong W, Pavasuthipaisit C, Santitadakul R. Validation of the Thai Version of the Patient Health Questionnaire for Adolescents (PHQ-A) in adolescent psychiatric patients. 2020;1(8):30–40.
30. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes. 2014;12:176.
31. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Appl Psychol Meas. 1977;1(3):385–401.
32. Cheng CP, Yen CF, Ko CH, Yen JY. Factor structure of the Center for Epidemiologic Studies Depression Scale in Taiwanese adolescents. Compr Psychiatry. 2012;53(3):299–307.
33. Lee SW, Stewart SM, Byrne BM, Wong JP, Ho SY, Lee PW, Lam TH. Factor structure of the Center for Epidemiological Studies Depression Scale in Hong Kong adolescents. J Pers Assess. 2008;90(2):175–84.
34. Wang M, Armour C, Wu Y, Ren F, Zhu X, Yao S. Factor structure of the CES-D and measurement invariance across gender in Mainland Chinese adolescents. J Clin Psychol. 2013;69(9):966–79.
35. Wood A, Kroll L, Moore A, Harrington R. Properties of the mood and feelings questionnaire in adolescent psychiatric outpatients: a research note. J Child Psychol Psychiatry. 1995;36(2):327–34.
36. Cao FL, Su LY, Cheng PX. Reliability and Validity of the Mood and Feelings Questionnaire in Chinese Adolescents (in Chinese). Chin J Clin Psychol. 2009;17:440–2.
37. Guy W. Clinical Global Impression Scale. ECDEU Assessment Manual for Psychopharmacology, Revised (DHEW Publ. No. ADM 76-338). Rockville: National Institute of Mental Health; 1976. p. 218–222.
38. Kadouri A, Corruble E, Falissard B. The improved Clinical Global Impression Scale (iCGI): development and validation in depression. BMC Psychiatry. 2007;7:7.
39. Bentler P, Bonett M, Douglas G. Significance tests and goodness of fit in the analysis of covariance structures. Psychol Bull. 1980;88(3):588–606.
40. Steiger JH. Structural Model Evaluation and Modification: An Interval Estimation Approach. Multivar Behav Res. 1990;25(2):173–80.
41. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociological Methods & Research. 1993;154(2):132–62.
42. Jöreskog KG, Sörbom D. Lisrel 8: Structural Equation Modeling With the Simplis Command Language; Lincolnwood: Scientific Software International, Inc; 1993.
43. Dancey CP, Reidy J: Statistics Without Maths for Psychology (7th edition); 2017.
44. Riedel M, et al. Response and remission criteria in major depression–a validation of current practice. J Psychiatr Res. 2010;44(15):1063–8.
45. Keller MB. Past, present, and future directions for defining optimal treatment outcome in depression: remission and beyond. JAMA. 2003;289(23):3152–60.

Publisher's Note