



BMJ Open Development and validation of a Chinese general practitioners' ability to recognise and manage depression assessment scale: a cross-sectional study

Yuqin Zhang ¹, Xuemei Sun,¹ Wen Zhong ², Dingkui Sun,³ Yingyan Gao,³ Xiyang Chen,⁴ Weiru Zhang^{1,2}

To cite: Zhang Y, Sun X, Zhong W, *et al.* Development and validation of a Chinese general practitioners' ability to recognise and manage depression assessment scale: a cross-sectional study. *BMJ Open* 2025;**15**:e094404. doi:10.1136/bmjopen-2024-094404

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-094404>).

Received 30 September 2024
Accepted 28 February 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

¹Department of General Practice, Xiangya Hospital, Central South University, Changsha, Hunan, China

²Xiangya Hospital Central South University National Clinical Research Center for Geriatric Disorders, Changsha, Hunan, China

³Central South University, Changsha, Hunan, China

⁴University of South China, Hengyang, China

Correspondence to
Dr Weiru Zhang;
zhangwr@csu.edu.cn

ABSTRACT

Objectives There is a shortage of reliable tools to evaluate general practitioners' (GPs) ability to recognise and manage depression. The main aim of this study was to develop and validate a reliable tool for assessing the GPs' ability to recognise and manage depression.

Design A cross-sectional survey incorporating the GPs' ability to recognise and manage the depression assessment scale, the work was administered between July 2023 and June 2024. GPs' ability to recognise and manage the depression assessment scale was developed in four phases: (1) item pool construction, (2) expert consultation, (3) exploratory factor analysis for further item reduction and to identify the factor structure of the revised scale and (4) confirmatory factor analyses to confirm the factors identified within the exploratory factor analysis.

Setting Primary healthcare in China.

Participants A total of 421 GPs participated in the study across phases 3 and 4. Phase 3 included 172 GPs, and phase 4 involved 249 GPs. They had been working in primary healthcare, with 44.42% having more than 10 years of experience. Of the participants, 57% were female.

Results The GPs' ability to recognise and manage the depression assessment scale comprised two dimensions, with 24 items. Two factors explained 66.62% of the items' variance through exploratory factor analysis. Confirmatory factor analysis confirmed two dimensions of the scale and produced appropriate Goodness of Fit Indexes. Reliability was robust, with composite reliability coefficients exceeding 0.60, Cronbach's alpha at 0.96 and a Spearman-Brown coefficient of 0.86.

Conclusions The newly developed scale is a reliable and valid tool for assessing GPs' abilities to recognise and manage depression. It is suitable for large-scale surveys, particularly in underdeveloped regions, and can help identify gaps in knowledge. The scale results highlight areas where GPs' skills are lacking, enabling the design of targeted continuing education programmes. It can also assess the effectiveness of depression training courses, providing a foundation for tailored interventions.

INTRODUCTION

Depression is one of the most prevalent mental health disorders globally, affecting individuals across all age groups and socioeconomic

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This study used a rigorous multiphase process, including expert consultation and psychometric validation to develop and validate the scale.
- ⇒ The scale's straightforward design makes it easy for general practitioners (GPs) to use, improving accessibility compared with traditional assessment tools.
- ⇒ The scale can be completed quickly and resource-efficiently, facilitating its widespread implementation.
- ⇒ Selection bias from purposive and snowball sampling limits the generalisability of the findings to all Chinese GPs.
- ⇒ Although the sample covers GPs across China, further validation in more diverse regions is needed to improve the scale's general applicability.

backgrounds.¹ Depression is associated with substantial personal, social and economic burdens that lead to decreased functioning, poorer quality of life, higher rates of medical comorbidities and increased morbidity and mortality.^{2,3}

In the context of primary care, general practitioners (GPs) serve as the first point of contact for many patients, making their role crucial in the early recognition and management of depression.⁴ It is estimated that approximately 60% of mental healthcare services are provided within primary care settings, and 79% of antidepressant prescriptions are issued by non-specialist healthcare providers.^{5,6} However, evidence suggests that many cases of depression go unrecognised or inadequately managed,⁷ particularly in primary healthcare.^{8,9} Multiple studies from both high-income and low- and middle-income countries (LMICs) have demonstrated that the recognition of depression in primary care remains inadequate. In high-income countries, more than half of

depression cases may go undetected,¹⁰ while the detection rates in LMICs are even lower.¹¹

This study aims to develop and validate a self-assessment scale that measures GPs' competence in recognising and managing depression. It allows GPs to assess their own skills in diagnosing and treating depression, helping them understand their strengths and areas that need improvement. Regular use of the tool helps GPs track their progress, identify areas where they need further training and support their career development. The main goal of the scale is to improve the quality of care that GPs provide to their patients by enabling more accurate recognition, timely treatments and better long-term management of depression. Additionally, the scale can be used for data collection and research. Policymakers and researchers can use it to gather information about GPs' competencies, which can help design training programmes and improve clinical practice. It can also serve as an evaluation tool for assessing the effectiveness of interventions like continuing education and training programmes. Ensuring that the scale is well-designed and validated is crucial to achieving these outcomes. By providing a reliable tool for self-evaluation, this study seeks to contribute to the enhancement of mental healthcare at the primary level and support GPs in addressing one of the most pressing public health challenges today.

METHODS

Participants' data remained confidential and anonymous, and all participants gave voluntary consent. The 'General Practitioners' Competency Scale for the Recognition and Management of Depression' was developed and validated following a multiphase methodology, as described by McCarthy *et al.*^{12 13} The detailed process is shown in figure 1.

Phase 1: item pool construction

A research panel consisting of one professor, six postgraduate students and four undergraduate students specialising in relevant fields developed the initial item pool. Before generating the initial item pool, we conducted a targeted literature review of depression and anxiety and clinical guidelines to identify core competencies and frequently encountered barriers. All panel members'

suggestions were documented and discussed in regular consensus meetings. We retained or modified items only after the majority of panel members (including the professor) reached an agreement, ensuring balanced input. The panel identified two key dimensions—recognition and management—based on clinical practice guidelines and expert consensus. Item pools for each dimension were then formulated through discussion.

Phase 2: expert consultation

Expert panel

A purposive sampling method was employed to recruit 20 experts with specialisations in general family medicine, mental health, psychiatry, public health and statistics.¹³ Each expert completed the consultation questionnaire independently, without group discussion or collaboration, to ensure unbiased and individual judgments. The expert panel was distinct from the GPs who participated in phases 3 and 4. The inclusion criteria were (1) a minimum of 5 years of experience in their respective fields, (2) holding a professional position at the associate advanced level or higher, (3) possessing substantial knowledge of depression and (4) providing voluntary consent to participate in the study.

Expert consultation form

From November 2023 to April 2024, this study collected expert opinions through a web-based survey (Wenjuanxin; <http://www.wjx.cn/>) and Word documents. The custom-made consultation questionnaire consisted of four parts: (1) an introduction explaining the study background, objectives and instructions for completing the survey; (2) demographic details of the experts, including gender, age, education and occupation; (3) a section for rating the importance of each item on a 5-point Likert scale (1=very unimportant, 5=very important), along with space for suggestions and additional items; and (4) a section on the experts' familiarity with the topic and their reasons for the judgments made.¹⁴ Further details are provided in online supplemental appendix A.

Following each round of expert consultation, the gathered data were carefully analysed. The research team reviewed the experts' suggestions regarding wording changes and item removal, then made necessary revisions and finalised decisions. A cut-off approach was employed, using three indicators: full score ratio, mean value and coefficient of variation (CV). The cut-off thresholds were determined as 'mean minus standard deviation (SD)' for the full score ratio and mean, and 'mean plus SD' for CV. Items that met at least two of these three criteria,¹⁵ along with any additional items suggested by experts, were retained for inclusion in the next round of consultation.

Further expert consultation was considered unnecessary when (1) no new feedback was provided by the experts, and (2) consensus was reached, with over 70% of the items in each dimension rated as 'important' or 'very important'.¹⁶ In total, three rounds of consultation were

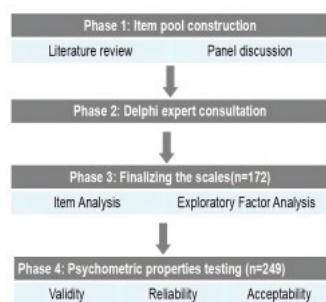


Figure 1 Procedure for the scale development and validation.

conducted, resulting in the creation of the initial versions of the assessment scales.

The reliability of experts

The reliability of expert evaluations was assessed using three indicators: (1) the expert positivity coefficient, determined by the response rate (%). A response rate above 70% was considered satisfactory; (2) the expert authority coefficient, calculated as the average of the experts' familiarity and judgement justification coefficients. A coefficient of 0.7 or higher was deemed acceptable; and (3) the consistency of expert opinions, represented by the CV for the importance ratings of items and Kendall's coefficient. Greater consistency in expert opinions is indicated by a lower CV and a higher Kendall's coefficient. Consistency was considered good if the CV was less than 0.25 and Kendall's coefficient exceeded 0.5 with a p value of less than 0.05.¹⁷

Phase 3: finalising the scales

The revised scales were finalised by administering them to a group of GPs. Item analysis and exploratory factor analysis (EFA) were used to evaluate the data and select the final items.

Sample and setting

EFA recommends a sample size that is 5–10 times the number of scale items.¹⁸ With 36 items in the revised scales, a sample size of over 150 participants was needed. Considering a 10% non-response rate, at least 165 participants were required. A snowball sampling method was used to survey 165 general practitioners in China, following the same inclusion criteria as the pilot survey.

Item pool revision

Participants were informed about the survey objectives and instructed to complete the scales anonymously, based on their ability to recognise and manage depression. Each item was rated on a five-point Likert scale: 'completely unable', 'occasionally able', 'able in half of the situations', 'mostly able' or 'completely able'. The total score was calculated, with higher scores reflecting a greater level of competence.

Data collection

The survey was conducted via an online questionnaire link (Wenjuanxin, <http://www.wjx.cn>) distributed to professional WeChat groups, academic forums and local healthcare networks from May to June 2024. Participants' responses and the scoring rules followed those used in the pilot survey. Surveys completed in under 60 s or those with identical responses for all items (eg, selecting the same option for every item) were excluded. These thresholds were determined based on pilot testing, which indicated that valid completion typically required at least 2–3 min for a 24-item scale. Additionally, GPs currently working in comprehensive hospitals were excluded, as the study focused on practitioners working primarily in community or primary care settings.

Item analysis

During item analysis, four indicators were calculated: (1) CV for each item, where items with a CV less than 0.20 were considered for removal¹⁷; (2) item discrimination: total scores were ranked, with the top 27% forming the high score group and the bottom 27% the low score group. This 27% threshold follows Kelley's (1939) recommendation for maximising group differences in item discrimination analysis, ensuring sensitivity to detect meaningful variations in competency levels.¹⁹ A t-test was used to compare these groups, as the sample size for each subgroup exceeded 30 (high-score group: n=46; low-score group: n=46), satisfying the central limit theorem assumption for normality. Prior studies have demonstrated the robustness of t-tests for large-sample comparisons without formal normality assessments.^{20,21} Items with no significant difference ($p>0.05$) were considered for deletion²²; (3) item-total correlation; and (4) Cronbach's α . The criteria for the last two indicators were the same as those used in the pilot survey.

Exploratory factor analysis

The assumptions for EFA were checked using the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity. A KMO value above 0.70 and a p value below 0.05 indicated the data were suitable for EFA. Principal axis factoring with maximum oblique rotation was used, and factors were extracted based on eigenvalues greater than 1.0. Items with factor loadings below 0.45 or communalities below 0.3 or if the item appeared in two or more factors at the same time (the difference in loadings is <0.1) were removed.^{23,24}

Phase 4: psychometric properties testing

To validate the final version of the scale, tests for construct validity, reliability and acceptability were conducted.

Sample, setting and data collection

The final version of the General Practitioners' Competency Scale for the Recognition and Management of Depression included 24 items, with 3 items removed during phase 3. The sample size for testing the psychometric properties followed EFA recommendations,²² resulting in a sample of 132 participants. The sampling method, setting and data collection procedures were consistent with phase 3, but this phase involved a new group of 172 GPs. No individual participated in more than one phase.

Construct validity

Construct validity was tested through model fit, convergent validity and discriminant validity. Confirmatory factor analysis (CFA) was used to evaluate model fit, with the following indices: χ^2/df , RMSEA (Root Mean Square Error of Approximation), CFI (Comparative Fit Index), GFI (Goodness-of-Fit Index), TLI (Tucker-Lewis Index) and SRMR (Standardized Root Mean Square Residual). A good model fit was indicated by $\chi^2/\text{df}<3.00$, $\text{RMSEA}\leq 0.08$, $\text{CFI}>0.90$, $\text{GFI}>0.90$, $\text{TLI}>0.90$ and $\text{SRMR}\leq 0.06$.^{25,26}

Convergent validity was assessed using average variance extracted (AVE) values, where values above 0.50 were considered good, 0.36–0.50 acceptable and below 0.36 poor. Discriminant validity was checked through correlation analysis, with good discriminant validity indicated by dimension–dimension correlations between 0.3 and 0.7, dimension–total score correlations between 0.7 and 1.0 and the square root of AVE being higher than the dimension–dimension correlations.²⁷

Reliability

Internal consistency reliability was assessed using the composite reliability coefficient, with values above 0.6 indicating good reliability.²⁸ Cronbach's α and split-half reliability were also used as supplementary measures. Split-half reliability was evaluated using the Spearman-Brown coefficient. Both Cronbach's α and Spearman-Brown coefficients ranged from 0 to 1, with values over 0.7 considered reliable.²⁹

Acceptability

An indicator used to assess the acceptability of the scales was completion time, referring to how long participants took to finish the scale. A completion rate above 85% and an average completion time of less than 20 min were considered signs of good acceptability.¹⁷

Statistical description and software

Categorical variables were reported as frequencies and percentages, while continuous data were presented as means with SDs or medians with IQRs, as appropriate. CFA was conducted using SPSSAU (www.SPSSA.com), and EFA along with other analyses were performed in SPSS V.25.0. A significance level of less than 0.05 was considered statistically significant.

Patient and public involvement statement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS

Phase 1: item pool construction

The item pool for the General Practitioners' Ability to Identify and Manage Depression assessment scale consisted of 24 items. Of these, 15 items focused on the recognition aspect, while 9 items addressed the management dimension. Detailed descriptions of the item pool content and the corresponding reference materials are provided in online supplemental appendix B.

Phase 2: expert consultation

20 experts, aged 32–60, from fields such as general family medicine, mental health, psychiatry and public health participated in this study. Detailed information on the experts is provided in online supplemental appendix C. A total of 20 questionnaires were sent out for the first round of expert consultation, all of which were returned with

valid responses. For the second and third rounds, 17 valid responses were collected. The response rates were 100% for the first round and 85% for the subsequent rounds, indicating strong expert engagement.

The expert consultation showed good authority and consistency, as evidenced by the expert authority coefficient, Kendall's coefficients and CV values for the item importance ratings meeting the required thresholds in the online supplemental appendix D. Throughout the consultation process, the following changes were made: (1) In the first round, 1 item was removed, 16 items were added and 22 items were revised. (2) In the second round, 9 items were deleted, 3 items were added and 6 items were revised. (3) In the third round, 6 items were deleted, with no further additions or revisions. The detailed results and reasons for item deletions are presented in online supplemental appendix E, F and G. After three rounds of consultation, the preliminary version of the General Practitioners' Ability to Identify and Manage Depression assessment, containing 27 items, was finalised.

Phase 3: finalising the scales

A total of 170 valid responses were received out of 172 distributed scales, resulting in a response rate of 98.84%. The participants included 60 males (35.29%) and 110 females (64.71%). Most participants (78.83%) had an undergraduate degree or higher, and more than half (55.29%) had over 10 years of work experience. The detailed characteristics of the participants are shown in online supplemental appendix H.

Item analysis and EFA

All items satisfied the criteria across the four indicators in the item analysis (table 1). The KMO value was 0.96, with a p value of less than 0.01, confirming the appropriateness of the data for conducting an EFA. Two common factors were identified, contributing a cumulative variance of 66.62%. The factor loadings of the assessment scale ranged from 0.46 to 0.61, while the communalities were between 0.60 and 0.83, all exceeding the threshold of 0.20 (table 2).

Scales finalisation

Three items were excluded due to the items appearing in two or more factors at the same time (the difference in loadings is <0.1): 'In clinical practice, the ability to distinguish depression symptoms from similar symptoms caused by organic conditions such as Parkinson's disease, Cushing's syndrome, hypothyroidism, autoimmune encephalitis, intracranial infections and stroke,' 'Considering the potential presence of depression when diagnosing and treating organic diseases (eg, coronary heart disease, diabetes, heart failure)' and 'The ability to inquire about suicidal intent and plans during clinical practice.' In one item, 'In clinical practice, be able to choose and correctly use appropriate screening scales (including PHQ-9 (Patient Health Questionnaire-9), BDI (Beck Depression Inventory), SDS (Self-rating Depression Scale),

Table 1 Item selection of the General Practitioners' Ability to Identify and Manage Depression assessment						
Dimensions	Item number	Critical ratio	CV*	Item-total correlation	Cronbach's alpha if item deleted†	Final conclusion
Recognition	1	7.501‡	0.16	0.600‡	0.974	Retention
	2	10.677‡	0.41	0.761‡	0.973	Retention
	3	11.249‡	0.42	0.742‡	0.974	Retention
	4	13.347‡	0.40	0.790‡	0.973	Retention
	5	12.249‡	0.40	0.793‡	0.973	Retention
	6	12.586‡	0.42	0.773‡	0.973	Retention
	7	15.494‡	0.43	0.795‡	0.973	Retention
	8	15.649‡	0.40	0.799‡	0.973	Retention
	9	12.843‡	0.33	0.775‡	0.973	Retention
	10	17.791‡	0.35	0.822‡	0.973	Retention
	11	12.666‡	0.43	0.735‡	0.974	Retention
	12	17.208‡	0.39	0.790‡	0.973	Retention
	13	16.972‡	0.38	0.794‡	0.973	Retention
Management	1	13.297‡	0.42	0.734‡	0.974	Retention
	2	11.688‡	0.38	0.773‡	0.973	Retention
	3	13.843‡	0.32	0.803‡	0.973	Retention
	4	13.340‡	0.34	0.776‡	0.973	Retention
	5	13.144‡	0.33	0.783‡	0.973	Retention
	6	13.204‡	0.34	0.797‡	0.973	Retention
	7	16.966‡	0.37	0.812‡	0.973	Retention
	8	13.372‡	0.32	0.830‡	0.973	Retention
	9	15.048‡	0.40	0.785‡	0.973	Retention
	10	13.229‡	0.37	0.820‡	0.973	Retention
	11	13.743‡	0.36	0.818‡	0.973	Retention
	12	12.145‡	0.44	0.673‡	0.974	Retention
	13	14.942‡	0.43	0.761‡	0.973	Retention
	14	15.740‡	0.31	0.819‡	0.973	Retention
*Coefficient of variation for the item score. †Overall Cronbach's alpha for the general practitioners' ability to recognise and manage depression assessment scale = 0.974. ‡p<0.01. CV, coefficient of variation.						

Table 2 The component matrix after rotation

Numbers of entries	Factors	
	Recognition	Management
A1	0.690	0.186
A2	0.821	0.304
A3	0.832	0.271
A4	0.793	0.370
A5	0.854	0.324
A6	0.852	0.297
A7	0.622	0.505
A8	0.620	0.511
A9	0.412	0.651
A10	0.571	0.582
A11	0.601	0.435
A12	0.587	0.523
A13	0.509	0.594
B1	0.380	0.613
B2	0.361	0.687
B3	0.230	0.851
B4	0.176	0.862
B5	0.279	0.781
B6	0.308	0.772
B7	0.405	0.710
B8	0.361	0.776
B9	0.427	0.651
B10	0.350	0.770
B11	0.357	0.761
B12	0.437	0.484
B13	0.321	0.794
B14	0.399	0.637

postpartum depression scales) for screening', although the loading does not fall within the original factor, the importance of group discussion has been recognised, and thus the item will be retained. Consequently, the finalised version of the assessment scale consists of 24 items—10 items focusing on recognition and 14 on management. A total score is calculated, where a higher score reflects greater recognition and management capabilities in the online supplemental appendix I.

Phase 4: psychometric properties testing

A total of 249 valid scales were returned out of 318, resulting in a rate of 78.3%. Among the participants, 121 (48.59%) were male, and 128 (51.41%) were female. The majority (78.83%) held an undergraduate degree or higher, and over half (55.29%) had more than 10 years of work experience. Detailed participant characteristics are presented in online supplemental appendix H.

Construct validity: The CFA results for the assessment scale indicated acceptable model fit, with $\chi^2/\text{df}=2.61$

(<3.00), RMSEA=0.08 (≤ 0.08), GFI=0.82 (>0.90), CFI=0.92 (>0.90) and SRMR=0.06 (<0.1). The AVE values exceeded 0.36, demonstrating acceptable convergent validity (0.58, 0.60). Additionally, the correlations between dimensions, as well as between each dimension and the total score, were within reasonable ranges. The square root of the AVE values surpassed the corresponding dimension–dimension correlation coefficients, indicating good discriminant validity for both scales (table 3).

Reliability

The reliability of the scale was robust, with composite reliability coefficients exceeding 0.60, Cronbach's alpha at 0.96 and a Spearman-Brown coefficient of 0.86. The Cronbach's α coefficients were 0.93 and 0.96, respectively, while the Spearman-Brown coefficients were 0.94 and 0.92, all above 0.70 (table 4), and composite reliability coefficients were 0.93 and 0.95, both exceeding the threshold of 0.60, indicating strong reliability for both scales.

Acceptability

The completion times ranged from 1 to 24 min (<30 min) with an average of 4 min, indicating good acceptability of the scale.

DISCUSSION

There is a lack of reliable assessment tools for GPs. To fill this gap, this study used a systematic approach to develop the 'General Practitioners' Ability to Identify and Manage Depression' scale. The scale has two dimensions: recognition and management, with a total of 24 items. Psychometric tests confirmed the scale's good construct validity, reliability and acceptability.

The scale's content is based on key resources, such as the mhGAP Intervention Guide, Depression in Primary Care: Part 1—Screening and Diagnosis, Part 2—Management and the Chinese Guidelines for the Diagnosis and Treatment of Depression in Primary Care (2021).^{30 31} Experts and researchers selected key clinical knowledge to form the item pool. Compared with existing guidelines, the scale is more practical and easier for GPs to use in daily practice. It provides a comprehensive assessment of GPs' ability to identify and manage depression.

Existing methods for GPs' ability to recognise depression often involve complex procedures, such as standardised patients or having psychiatrists re-evaluate patients diagnosed by GPs.^{7 32} While these assessments often yield highly accurate evaluations of clinical performance, they also require substantial time, specialised personnel and technical resources. In addition, although formal exams can assess knowledge, they may not capture the day-to-day clinical reasoning involved in real-world practice. Compared with the Depression Attitude Questionnaire, our scale focuses on practical competence (recognition and management). In contrast, the self-assessment tool developed in this study offers a more straightforward and

Table 3 Correlation coefficients and average variance extracted square root values of the assessment scale

Categories	Recognition	Management
Recognition	0.762	
Management	0.740	0.774

The diagonal is the average variance extracted square root values. *p value <0.01.

cost-effective alternative, making it especially suitable for widespread use in primary care settings with limited resources. However, this study did not include a direct comparison with other scales, as our primary aim was to develop and validate a reliable tool specifically tailored to assess GPs' ability to recognise and manage depression. Future research will focus on comparing our scale with existing tools.

Moreover, the self-assessment tool can serve as a valuable guide for continuing education and training programmes. As usage of this assessment scale expands, a structured feedback system can provide GPs with individualised reports, highlighting strengths and flagging specific areas for improvement (eg, screening scale, physical symptoms, differential diagnosis). Professional bodies and academic institutions could then develop targeted workshops or online modules addressing these identified gaps. On a broader level, aggregated, anonymised results may guide system-wide training initiatives, ensuring resources are directed where they are most needed. Periodic re-assessment would enable ongoing quality improvement, monitoring the effectiveness of these educational interventions and continually refining the approach to depression management in primary care.

This study successfully developed and validated the 'General Practitioners' Ability to Identify and Manage Depression' scale, but some limitations should be noted. The use of purposive and snowball sampling methods may introduce selection bias, limiting the generalisability of the findings. Despite efforts to include participants from diverse regions (East, Central and West China), the sample may not fully represent all Chinese GPs. The sample size used in our EFA, although meeting the basic thresholds, was near the lower bound of commonly recommended guidelines. Kendall's coefficient of 0.211 and a CV of 0.17 indicated satisfactory expert consistency in the Delphi process. The study focused on scale development and initial validation, with further research needed on the subgroup differences. Future studies should conduct external and prospective validation in broader settings to ensure generalisability.

Good psychometric properties are essential to ensure the accuracy and usefulness of the scale.²¹ Since the scale is designed for GPs, it explains specific competencies in simple terms, making it easier for doctors to assess their abilities. The scale has 24 items and takes about 4 min to complete. It is easy to administer and requires minimal resources. Therefore, it is a reliable, valid, time-saving tool that can be widely used in Chinese primary health-care settings.

Implications for future research and practice

Developing a scale to measure general practitioners' ability to recognise and manage depression is a crucial step in improving mental healthcare in primary care settings. Assessing knowledge and skills in this area will enable the identification of gaps and inform targeted training programmes aimed at enhancing the capabilities of general practitioners. This will lead to better patient outcomes and a more efficient use of healthcare resources. Clinicians and educators are encouraged to use the scale to assess general practitioners' competencies and to tailor educational interventions based on the results. Additionally, the scale can serve as a valuable tool for evaluating the effectiveness of these interventions. Future research may focus on translating and validating the scale across different languages and healthcare systems, facilitating broader collaboration and advancing mental healthcare globally.

CONCLUSION

The GPs' ability to recognise and manage the depression assessment scale was successfully developed and demonstrated to be reliable, valid, easy to apply, time-efficient and resource-efficient for assessing the ability of general practitioners in managing depression. This scale can be used to promptly identify ability gaps, which will guide the development of targeted educational interventions and allow for the evaluation of their effectiveness. Furthermore, it serves as a useful tool for improving general

Table 4 Results of the reliability evaluation of scales

	Cronbach's α	Split-half reliability	Composite reliability
Recognition	0.93	0.94	0.93
Management	0.96	0.92	0.95

practitioners' capabilities in depression management, ultimately enhancing patient care outcomes.

Contributors WRZ is the guarantor. Conceptualisation: WZ and YYG. Data curation: XMS, DKS and XYC. Normal analysis: XMS, DKS and YQZ. Investigation: XMS and YQZ. Supervision: WRZ. Writing original draft: YQZ. Writing, reviewing and editing: YQZ, XMS, WZ, DKS, YYG, XYC and WRZ.

Funding This work was supported by the Natural Science Foundation of Hunan Province (No. 2022JJ40846)

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The Ethics Committee of Xiangya Hospital, Central South University, China (2024101215). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. Data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Yiqin Zhang <http://orcid.org/0009-0001-5897-353X>

Wen Zhong <http://orcid.org/0000-0002-2444-3047>

REFERENCES

- McCarron RM, Shapiro B, Rawles J, *et al.* Depression. *Ann Intern Med* 2021;174:ITC65–80.
- Cuijpers P, Quero S, Dowrick C, *et al.* Psychological Treatment of Depression in Primary Care: Recent Developments. *Curr Psychiatry Rep* 2019;21:129.
- Malhi GS, Mann JJ. Depression. *Lancet* 2018;392:2299–312.
- Park LT, Zarate CA. Depression in the Primary Care Setting. *N Engl J Med* 2019;380:559–68.
- Frank RG, Huskamp HA, Pincus HA. Aligning incentives in the treatment of depression in primary care with evidence-based practice. *Psychiatr Serv* 2003;54:682–7.
- Mark TL, Levit KR, Buck JA. Datapoints: psychotropic drug prescriptions by medical specialty. *Psychiatr Serv* 2009;60:1167.
- Kirmayer LJ, Robbins JM, Dworkind M, *et al.* Somatization and the recognition of depression and anxiety in primary care. *Am J Psychiatry* 1993;150:734–41.
- Montano CB. Recognition and treatment of depression in a primary care setting. *J Clin Psychiatry* 1994;55:18–34.
- Habtamu K, Birhane R, Demissie M, *et al.* Interventions to improve the detection of depression in primary healthcare: systematic review. *Syst Rev* 2023;12:25.
- Gilbody S, Sheldon T, House A. Screening and case-finding instruments for depression: a meta-analysis. *CMAJ* 2008;178:997–1003.
- Udedi M. The prevalence of depression among patients and its detection by primary health care workers at Matawale Health Centre (Zomba). *Malawi Med J* 2014;26:34–7.
- McCarthy GS, Truxillo DM, O'Shea DE, *et al.* The development and validation of a Multidimensional Perceived Work Ability Scale. *J Occup Health Psychol* 2024;29:90–112.
- McPherson S, Reese C, Wendler MC. Methodology Update: Delphi Studies. *Nurs Res* 2018;67:404–10.
- Li Y, Ehiri J, Hu D, *et al.* Framework of behavioral indicators for outcome evaluation of TB health promotion: a Delphi study of TB suspects and Tb patients. *BMC Infect Dis* 2014;14:268.
- Hou YF, Liu D, Zhou SG, *et al.* The capacity for psychotherapy process scale: development and psychometric validation of patient self-reports. *Psychother Res* 2021;31:236–46.
- de Villiers MR, de Villiers PJT, Kent AP. The Delphi technique in health sciences education research. *Med Teach* 2005;27:639–43.
- Sun X, Gao Y, Chen Y, *et al.* Development and validation of frailty and malnutrition knowledge assessment scale for community-dwelling older adults. *Appl Physiol Nutr Metab* 2023;48:974–1004.
- Wu ML. *SPSS statistical applications: questionnaire analysis and applications*. Beijing: Beijing Science Press, 2003.
- Kelley TL. The selection of upper and lower groups for the validation of test items. *J Educ Psychol* 1939;30:17–24.
- Lumley T, Diehr P, Emerson S, *et al.* The importance of the normality assumption in large public health data sets. *Annu Rev Public Health* 2002;23:151–69.
- Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008;65:2276–84.
- Boateng GO, Neilands TB, Frongillo EA, *et al.* Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front Public Health* 2018;6:149.
- Wang H, Gai X, Li S. Development and Validation of the Youth Purpose Orientation Scale Among Chinese Sample. *Psychol Res Behav Manag* 2022;15:3485–96.
- Y: MM. Revision and application of the scale of knowledge, attitude and ability to deal with elder abuse of medical staff. 2022.
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1–55.
- Hu L, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol Methods* 1998;3:424–53.
- Li J, Zhang L, Zhou J, *et al.* Development and validation of self-management scale for tuberculosis patients. *BMC Infect Dis* 2022;22:502.
- Wei J, Guo L, Zhang JF. The effect of representation effect on the internal consistency reliability of balanced scales. *Southwest Univ* 2017.
- Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011;2:53–5.
- Ferenchick EK, Ramanuj P, Pincus HA. Depression in primary care: part 1-screening and diagnosis. *BMJ* 2019;365:l794.
- Ramanuj P, Ferenchick EK, Pincus HA. Depression in primary care: part 2-management. *BMJ* 2019;365:l835.
- Fekadu A, Medhin G, Selamu M, *et al.* Recognition of depression by primary care clinicians in rural Ethiopia. *BMC Fam Pract* 2017;18:56.