ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202510.00115

The Heterogeneity of Youth Depression Scales

Authors: Wang,Haoyuan, Hu,Mengzhen, Tian, Liuqing, Liu, Weibiao, An,Yuanyuan, Li,Ying, Chuanpeng Hu, Li, Ying, Hu,Chuanpeng

Date: 2025-12-08T11:46:32+00:00

Abstract

Depression is a leading mental health concern among children and adolescents,—90% of whom live in developing countries. Accurate measurement of depression in this population is a crucial step toward effective research, diagnosis, and intervention. Although rating scales for children and adolescents are widely used by researchers, clinicians, and large-scale surveys, their heterogeneity in developing contexts has been largely overlooked. To fill the gap, we analyzed the content of 27 Chinese depression scales and self-report data from four commonly used scales (N = 12,764). Our content analysis identified 84 unique symptoms across the 27 scales and low rate of overlap (mean = 0.19, range: 0.09–0.25), suggesting that instruments often assumed to becomparable capture distinct symptom domains. Moreover, few scales included culturalor age-specific symptoms, suggesting a lack of cultural- or age-adaptation. Analyses of self-report data showed that the four scales produced different detection rates and each identified unique cases. Exploratory Graph Analysis revealed that their items clustered into three distinct symptom communities rather than a unified construct. These findings indicate that depression scales for children and adolescents are not interchangeable and highlight the need for harmonization efforts. Achieving a balance between standardization and contextual adaptation is essential for improving the validity and comparability of depression measurement in global mental health research.

Full Text

Preamble

The Heterogeneity of Youth Depression Scales

Haoyuan Wang[1], Mengzhen Hu[1], Liuqing Tian[1], Weibiao Liu[1], Yuanyuan An[1], Ying Li[2]#, Hu Chuan-Peng[1]#

[1]School of Psychology, Nanjing Normal University, Nanjing, China
[2]Department of Psychosomatic Medicine, Beijing Children' s Hospital, Capital Medical University, Beijing, China

CrediT Author Statement
Haoyuan Wang: Conceptualization, Data curation, Visualization, Investigation, Writing -Original Draft, Writing -Reviewing and Editing.
Mengzhen Hu: Data curation, Visualization, Investigation, Writing -Reviewing and Editing.
Liuqing Tian: Investigation, Writing -Reviewing and Editing.
Weibiao Liu: Investigation, Writing -Reviewing and Editing.
Yuanyuan An: Writing -Reviewing and Editing.
Ying Li: Investigation, Writing -Reviewing and Editing, Data curation.
Hu Chuan-Peng: Conceptualization, Supervision, Investigation, Project administration, Writing -Original Draft, Writing -Reviewing and Editing.

Corresponding authors:
Hu Chuan-Peng, email: hcp4715@hotmail.com
Ying Li, email: liying@bch.com.cn

Abstract

Depression is a leading mental health concern among children and adolescents—90% of whom live in developing countries. Accurate measurement of depression in this population is a crucial step toward effective research, diagnosis, and intervention. Although rating scales for children and adolescents are widely used by researchers, clinicians, and large-scale surveys, their heterogeneity in developing contexts has been largely overlooked. To fill this gap, we analyzed the content of 27 Chinese depression scales and self-report data from four commonly used scales (N = 12,225). Our content analysis identified 84 unique symptoms across the 27 scales and a low rate of overlap (mean = 0.19, range: 0.09–0.25), suggesting that instruments often assumed to be comparable capture distinct symptom domains. Moreover, few scales included cultural- or age-specific symptoms, suggesting a lack of cultural- or age-adaptation. Analyses of self-report data showed that the four scales produced different detection rates and each identified unique cases. Exploratory Graph Analysis revealed that their items clustered into three distinct symptom communities rather than a unified construct. These findings indicate that depression scales for children and adolescents are not interchangeable and highlight the need for harmonization efforts. Achieving a balance between standardization and contextual adaptation is essential for improving the validity and comparability of depression measurement in global mental health research.

Keywords: Depression; Youth; Heterogeneity; Measurement; Non-Western Culture

Introduction

Depression is one of the most common mental health problems among children and adolescents: about one in five currently experience depression or report depressive symptoms [?, ?]. Notably, 90% of children and adolescents live in developing countries [?, ?]. Moreover, the prevalence of depression among children and adolescents in developing countries has been rising rapidly [?, ?]. As convenient and low-cost tools for assessing the presence and severity of depression, rating scales are widely used for screening depression in children and adolescents [?, ?], especially in middle- and low-income regions [?, ?]. Rating scales not only play a pivotal role in depression screening and diagnosis but also serve as a foundation for understanding and intervening in depression [?, ?]. For example, in clinical trials, researchers have relied heavily on the Children' s Depression Rating Scale-Revised (CDRS-R) to evaluate the efficacy of antidepressants in children and adolescents [?, ?]. An individual' s decisions might be changed by the testing scores of depression scales: adolescents who self-identify with depressive symptoms through online platforms may develop heightened negative thinking and interpersonal withdrawal [?, ?]. When used at a larger scale, scales may change public policy [?, ?]. In China, for example, the increased prevalence of depression, measured primarily by rating scales, has resulted in more new policies that aim to alleviate the issue [?, ?]. Globally, the Measurement of Mental Health among Adolescents at the Population Level (MMAP) initiative, initiated by UNICEF and WHO, also relies on standardized rating scales and aims to provide the evidence for mental health policies among developing countries [?, ?].

Despite the central role of rating scales for depression among children and adolescents, the heterogeneity of these tools has often been overlooked. Previous studies have revealed significantly different detection rates measured by different tools. For example, based on structured clinical interviews, the prevalence of major depressive disorder among Chinese children aged 6 to 16 is estimated at 2% to 3% [?, ?, ?]. However, self-report rating scales yield substantially higher detection rates, ranging from 6% to 17% in primary school students, 13% to 38% in middle school students, 5.2% to 42.3% in high school students, and 3.9% to 45.9% in college students [?, ?, ?, ?, ?]. These discrepancies suggest that rating scales vary considerably in what they are actually measuring. This pattern echoes the heterogeneity among scales of depression for adults. For instance, Fried (2017) found low overlap across seven common depression scales. Similarly, Veal et al. (2024) identified 388 distinct instruments across 450 randomized controlled trials for unipolar and bipolar depression. These findings underscore the importance of understanding how depression is conceptualized and measured among children and adolescents across the world, given that depression among children varies across different social contexts and differs from adolescents [?, ?]. However, no such systematic investigations have yet been conducted among child and adolescent populations in developing countries.

To fill this gap, we systematically analyzed symptoms measured by 27 Chinese

scales and quantified the differences between scales by re-analyzing data from 12,225 children and adolescents who had taken four commonly used depression scales. To our knowledge, this is the first systematic evaluation of depression rating scales for children and adolescents conducted outside of Western cultural contexts. The findings confirm the heterogeneity previously observed in adult-focused studies and reveal broader methodological concerns, including the over-reliance on Western-developed tools [**?**, **?**, **?**] and the neglect of culture- and age-specific symptomatology [**?**, **?**]. Our empirical data further confirmed that four commonly used rating scales are not measuring a unidimensional latent construct. Together, our findings call for better depression assessment practices in non-Western youth populations.

Study 1: Content Analysis

Methods

Study Procedure

We took three steps to code symptoms from all scales that measure depression in children and adolescents (i.e., students from elementary school to college). Firstly, we identified all scales that had been used for screening depression from four comprehensive meta-analyses. Secondly, we identified unique symptoms within each scale. Thirdly, we compared the symptoms and identified unique symptoms across scales. The latter two steps followed Fried (2017) with minor modifications (see Supplementary Methods S1.1 and Fig. S1 for details). Inter-rater agreement was assessed to ensure the robustness of the coding procedure, using Cohen' s [**?**, **?**].

Identifying depression scales

We identified scales that measure depression in children and adolescents from four recent meta-analyses, which synthesized the prevalence of different mental health problems among four Chinese student populations: elementary school, middle school, high school, and college [**?**, **?**, **?**, **?**, **?**]. Across 441 depression-related articles in these four meta-analyses, we found 33 unique scales. We then screened the versions of scales and selected the most suitable version for later analysis (see Supplementary Methods S1.2 for details).

Identify unique symptoms within scales

After identifying all scales with a specific version, four trained coders assessed identical or similar symptoms within each scale. First, the four coders independently identified items that assess the same symptom within each scale (see Supplementary Methods S1.3 and Supplementary Table S1 for details). Then, the four coders teamed up as two pairs, and each pair reviewed their members' results and resolved any discrepancies within the pair. Subsequently, the results from the two pairs of coders were cross-checked, and any inconsistencies were discussed and resolved with the corresponding team members. The merged items were verified independently by a clinically certified psychiatrist (Y. L).

Identify unique symptoms across scales

In this step, four coders compared symptoms across all scales and merged items that assess the same symptoms. The procedure was the same as identifying unique symptoms within scales: independent individual coding, discussion by pairs, cross-checking between pairs, discussion with the corresponding author, and verification by a clinically certified doctor.

When identifying unique symptoms across scales, we retained both compound symptoms and specific symptoms, as in Fried (2017). Compound symptoms are symptoms that include a range of related symptoms, whereas specific symptoms are more concrete and describe specific patterns. For example, "appetite changes" is a compound symptom; it includes two specific symptoms: "appetite increased" and "appetite decreased," and all three of them were treated as a unique symptom. We employed an approach that maximized the number of different symptoms. More specifically, if the items describe similar symptoms using different words and the words have significantly different meanings under the Chinese context, we treat them as belonging to the same compound symptom but as different specific symptoms. For instance, there are many different words to describe depressed mood in different scales; we used 'depressed moods' as the compound symptom but distinguished different specific symptoms such as 'blue', 'low mood', 'sad', and 'anhedonia'. This approach is slightly different from Fried (2017), where he coded all these items as a specific symptom 'Sad moods' (see Supplementary Table S2 for details).

After identifying the unique symptoms, we assigned scores for all scales on all unique symptoms. More specifically, a scale was scored as "0" on a symptom if it does not have items that measure this symptom. For instance, the Children's Depression Inventory (CDI) has no item for 'Depressed mood', so we assigned "0" for CDI on this symptom. If a scale has an item that directly measures a symptom—compound or specific—it was coded as 2 on that symptom. Note that if a scale has an item measuring a compound symptom, then this scale not only had a score of 2 on that compound symptom but also had a score of 1 on all specific symptoms of this compound symptom. For example, CDI has an item that directly measures the compound symptom "appetite change" and scored 2 on this compound symptom. Importantly, even though the CDI does not have items for 'appetite increased' and 'appetite decreased', it scored 1 on these two specific symptoms (see https://osf.io/2s5dz for details). However, if the item measures a specific symptom under a compound symptom, this scale was coded "2" on that specific symptom but still "0" on the corresponding compound symptom.

Data analyses

We reported the descriptive summary of scales as well as the symptoms within each scale. We highlighted symptoms that are used in DSM-5 for diagnosis of depression. More specifically, there are 28 symptoms overlapped with the DSM-5 symptoms for depression, which were derived from the nine symptoms in DSM-5 and all their specific symptoms [**?**, **?**].

We used the Jaccard index to quantify the degree of overlap between different

scales [**?**, **?**]. The formula of Jaccard index is s/(u1 + u2 + s), where "s"represents the number of items shared by two scales, and "u1"and "u2"denote the number of items that are exclusively present in each of the two scales. Jaccard index ranges from 0 (no overlap among scales) to 1 (complete overlap). We interpreted the Jaccard index as in Fried (2017): very weak 0.00–0.19, weak 0.20–0.39, moderate 0.40–0.59, strong 0.60–0.79, and very strong 0.80–1.0. Moreover, we explored the relationship between the mean Jaccard coefficient, adjusted scale length, and the number of captured symptoms (i.e., number of symptoms included in the scale) by employing Spearman correlation.

Results

The inter-rater agreements during the coding phase reached substantial to nearly perfect agreement across coders (= 0.71–0.96, $p < .001$; 0.60–0.79 = substantial, 0.80–0.99 = nearly perfect; [**?**, **?**]; see Supplementary Methods S1.1).

We summarized the number of articles using each depression scale from four meta-analyses and plotted the relationship between usage frequency and the number of symptoms captured (see Fig. S2).

We identified 385 items from 27 scales identified in four recent meta-analyses (see Methods and Supplementary Results S1.2 for details). After comparing within each scale as well as between scales, items were merged into 84 unique symptoms (see Fig. 1 [Figure 1: see original paper]), including 8 compound symptoms (Depressive mood, Irritability, Self-abasement, Interest/pleasure loss, Somatization, Appetite changes, Somnipathy, and Reduced socialization).

There were 18 (21.43%) idiosyncratic symptoms. Nineteen scales out of twenty-seven did not include any idiosyncratic symptoms. For the other eight scales, the rate of idiosyncratic symptoms varied from 3.9% to 22.2%. Interestingly, only 8 scales (29.62%) included items specific for children and adolescents; these items were merged as 24 unique symptoms (28.57% of all unique items), as detailed in Supplementary Table S3.

No single symptom was present in all scales. The most frequent symptom, appearing in 22 out of 27 scales, was sense of hopelessness. The second was interest loss, which appeared in 18 out of 27. Note that markedly diminished interest or pleasure, a key symptom for the diagnosis of major depression in DSM-5, is split into two specific symptoms: interest loss and pleasure loss in this study. We found pleasure loss was observed less frequently than interest loss, being measured in 9 out of 27 scales.

Another frequently measured symptom is the compound symptom Depressed mood, which was directly measured in 5 scales. However, this compound symptom includes several specific symptoms: blue appeared in 10 scales, low mood in 15 scales, sad in 13 scales, and anhedonia in 16 scales. Combined as a cluster, depressed mood and its related specific symptoms were present in 25 out of 27 scales and were the most frequent symptoms.

Interestingly, none of the scales covered all the DSM-5 symptoms. The Depression Status Inventory (DSI) led the coverage ranking with a rate of 71.42%. Please see Supplementary Table S3 for detailed information.

The degree of overlap between scales was calculated using the Jaccard coefficient. The mean overlap across all scales was 0.19, ranging from 0.09 to 0.25, indicating a very low level of similarity between these scales (see Fig. 2 [Figure 2: see original paper]). The Center for Epidemiological Studies Depression Scale (CES-D) has the highest mean degree of overlap with other scales. The highest overlap, 0.75, appeared between two versions of CES-D: CES-D for adults and CES-D-C for children. The second highest overlap, 0.72, was between Depression Status Inventory and Self-Rating Depression Scale. Many scales had no overlap at all. For example, there was no overlap between China Education Panel Survey (CEPS) and the Middle-school students Mental Health Scale (MSSMHS), Patient Health Questionnaire-9 items (PHQ-9), China Middle school students' Depression Scale (CSSDS), Short Mood and Feelings Questionnaire (SMFQ), Chinese College Student Mental Health Scale (CCSMHS). Note that because Comprehensive Survey Report on Health-Related/Risk Behaviors among Chinese Adolescents (Ji_{2007}) has only one item that measures two symptoms, it has no overlap with Patient Health Questionnaire-9 items (PHQ-9), Kutcher Adolescent Depression Scale (KADS-11). The mean overlap of each scale was correlated with the length of the scale, r = 0.70, 95% CI [0.42, 0.87]. The mean overlap of scales was correlated with the number of captured symptoms (r = 0.54, 95% CI [0.17, 0.78]).

Figure 1. Content Overlap Across Twenty-seven Depression Scales. Each row represents a symptom, each column represents a scale. If a scale measures a symptom, then there is a dot or a circle on that row. The former represents the scale directly captures symptom and the latter represents the scale indirectly captures symptom. There are 18 symptoms that appear only in one scale; these symptoms are referred to as idiosyncratic symptoms. Red indicates age-specific scales or symptoms, while blue indicates culturally specific scales or symptoms. ADI: Adolescent Depression Inventory; CDI: Children' s Depression Inventory; HAMD: Hamilton Depression Rating Scale for Depression; DSI: Depression Status Inventory; SDS: Self-Rating Depression Scale; MFQ-C: Mood and Feelings Questionnaire; CBCL: Child Behavior Checklist; BDI-II: Beck Depression Inventory-II; DSRSC: Depression Self-rating Scale for Children; BDI-I: Beck Depression Inventory; KADS-11: Kutcher Adolescent Depression Scale; CES-D: The Center for Epidemiological Studies Depression Scale; PHQ-9: Patient Health Questionnaire-9 items; CSSDS: China Middle school students' depression scale; CES-D-C: Center for Epidemiologic Studies Depression Scale for Children; UPI: University Personality Inventory; SMFQ: Short Mood and Feelings Questionnaire; SCL-90: Symptom Checklist 90; CES-D-13: Short version of Center for Epidemiologic Studies Depression Scale; CCSMHS: Chinese College Student Mental Health Scale; DASS-21: The Depression Anxiety Stress Scale-21 Items; BSRS: Brief Symptom Rating Scale; Sakuma_{2010}: Sakuma et al. (2010) self-designed scale; MSSMHS: Middle-school students

Mental Health Scale; CEPS: China Education Panel Survey; HADS: Hospital Anxiety and Depression Scale; Ji_{2007}: Comprehensive Survey Report on Health-Related/Risk Behaviors among Chinese Adolescents.

Figure 2. Overlap of item content of 27 depression scales. ADI: Adolescent Depression Inventory, CDI: Children' s Depression Inventory, HAMD: Hamilton Depression Rating Scale for Depression, DSI: Depression Status Inventory, SDS: Self-Rating Depression Scale, MFQ-C: Mood and Feelings Questionnaire, CBCL: Child Behavior Checklist, BDI-II: Beck Depression Inventory-II, DSRSC: Depression Self-rating Scale for Children, BDI-I: Beck Depression Inventory, KADS-11: Kutcher Adolescent Depression Scale, CES-D: The Center for Epidemiological Studies Depression Scale, PHQ-9: Patient Health Questionnaire-9 items, CSSDS: China Middle school students' depression scale, CES-D-C: Center for Epidemiologic Studies Depression Scale for Children, UPI: University Personality Inventory, SMFQ: Short Mood and Feelings Questionnaire, SCL-90: Symptom Checklist 90, CES-D-13: Short version of Center for Epidemiologic Studies Depression Scale, CCSMHS: Chinese College Student Mental Health Scale, DASS-21: The Depression Anxiety Stress Scale-21 Items, BSRS: Brief Symptom Rating Scale, Sakuma_{2010}: Sakuma et al. (2010) self-designed questionnaire, MSSMHS: Middle-school students Mental Health Scale, CEPS: China Education Panel Survey, HADS: Hospital Anxiety and Depression Scale, Ji_{2007}: Comprehensive Survey Report on Health-Related/Risk Behaviors among Chinese Adolescents. Mean overlap is detailed in Supplementary Table S3.

Study 2: Empirical Research

Methods

We re-analyzed a dataset from the Department of Psychosomatic Medicine at Beijing Children' s Hospital from May 2020 to August 12, 2024. Out of 12,290 total participants who visited the Department for the first time, we selected data for those aged 7 to 18, yielding 12,225 participants, including 6,468 males and 5,757 females. All participants' parents agreed to participate in the study and signed informed consent forms. The study was approved by the IRB at the hospital.

Inclusion Criteria (Participants must meet ALL of the following): - Age: 7–18 years - Consent: Legal guardian(s) provided signed informed consent - Initial assessment: Completed all self-report scales at the first clinical visit

Exclusion Criteria (Participants are excluded if they meet ANY of the following): - Severe Comorbidity: Neurological disorders (e.g., epilepsy, traumatic brain injury); Neurodevelopmental disorders (e.g., Autism Spectrum Disorder, intellectual disability) - Active psychosis or substance abuse - Incomplete Data: Missing critical diagnosis or symptom records

All included participants completed four self-report depression scales in the fol-

lowing fixed order: Chinese version of Depression Self-Rating Scale for Children, Chinese version of Patient Health Questionnaire-9 items, Chinese version of The Depression Anxiety Stress Scale-21 Items, and Chinese version of Children' s Depression Inventory.

Depression scales

Four depression scales were used for measuring depression in this dataset. All of them were included in our Study 1.

The Chinese version of Depression Self-Rating Scale for Children (DSRSC) [**?**, **?**] has 18 items. Participants rated their symptoms over the past week on a 0-2 scale: Not at all (0), Sometimes (1), and Often (2). The sum score was used and higher scores indicated more severe depression. Participants with sum score higher than 15 are regarded as depressed [**?**, **?**].

The Chinese version of Children' s Depression Inventory (CDI) consists of 27 items [**?**, **?**]. Participants rate their symptoms over the past two weeks from 0 to 2, indicating different levels of depression: mild, moderate, and severe (e.g., "I feel sad sometimes," "I often feel sad," "I feel sad all the time"). The sum scores vary from 0 to 54, with higher sum scores indicating more severe depression. Item 26 is not scored, and the cutoff score is 20 [**?**, **?**].

The Chinese version Patient Health Questionnaire (PHQ-9) comprises 9 items [**?**, **?**]. Participants reported how frequently they experienced each symptom over the past two weeks by selecting one of four options: "Not at all" (0), "Several days" (1), "More than half of all the days" (2), and "Nearly every day" (3). A sum score of 10 or higher is used as the cutoff for screening [**?**, **?**].

The depression subscale of the Chinese version of Depression Anxiety Stress Scale (DASS-21) consists of 7 items [**?**, **?**]. Participants were presented statements that describe experiences over the past week (e.g., "I feel depressed," "I feel dry in my mouth") and were asked to rate how each item applied to themselves on a 4-point scale: "never" (0), "sometimes" (1), "often" (2), and "almost always" (3). Sum scores of these 7 items were used and the cutoffs are 4 for males and 5 for females [**?**, **?**].

All these four scales were analyzed in Study 1. Their 61 items were combined into 56 unique items during the within-scale coding (see Supplementary Table S1 and https://osf.io/vt2xg for details) and further combined into 39 items that measure 40 unique symptoms across scales (see https://osf.io/vt2xg). Note that item 8 of PHQ-9 simultaneously measures two symptoms, Retardation and Agitation. The symptoms measured by these four scales are shown in Fig. 1 and on https://osf.io/vt2xg. The overlap between these four scales ranges from 0.04 to 0.26 (see Fig. 2 and Supplementary Table S4 for details).

Data Analyses

To test the heterogeneity of the four scales, we first compared the detection rates of depression across different scales based on the criteria of each scale. The differences in detection rates indicate differences of the scales in diagnosis

of depression. We also tested whether the symptoms in these four scales can form a single community by using Exploratory Graph Analysis [**?**, **?**]. Exploratory Graph Analysis is a novel method in network psychometrics for identifying dimensions in multivariate data using network models. In these models, nodes (circles) represent variables, while edges (lines) reflect the relationships (e.g., partial correlations) between the nodes. Communities of nodes are interpreted as dimensions within the network [**?**, **?**]. EGA is based on the estimation of a network followed by the application of a community detection algorithm [**?**, **?**], which works well in both unidimensional and multidimensional structures [**?**, **?**]. We used the EGAnet package in R for the EGA analyses [**?**, **?**] and followed the standard workflow [**?**, **?**]: (1) determining redundancies of input variables, (2) performing EGA, and (3) checking the stability of EGA using bootEGA (see Supplementary Methods S3.2). If all these symptoms are measuring the same latent variable, we should expect a unidimensional structure of these symptoms; otherwise, these symptoms are measuring different "latent variables."

All code is available at: https://github.com/Chuan-Peng-Lab/YouthDepressioninScales

Results

To further validate the heterogeneity of depression scales, we analyzed an existing large dataset with 12,225 children and adolescents in the hospital who answered four commonly used depression scales: the Depression Anxiety Stress Scale-21 Items (DASS-21), Patient Health Questionnaire-9 items (PHQ-9), Depression Self-rating Scale for Children (DSRSC), and Children' s Depression Inventory (CDI) (see Methods for details). These four scales covered 39 symptoms of all 84 symptoms we identified among 27 scales. We tested the following hypotheses: (1) If these scales are heterogeneous, then their detection rates of depression would differ; (2) If symptoms covered by these scales are heterogeneous, then symptoms would form multiple communities/clusters instead of a single one.

Our results revealed that detection rates varied across scales for the same sample: DASS (62.43%), PHQ-9 (53.46%), DSRSC (52.03%), and CDI (49.25%). This pattern was the same for different age groups (see Fig. 3A [Figure 3: see original paper]). Furthermore, each scale detected some unique cases that were not detected by other scales: DASS flagged 833 unique cases, PHQ-9 326, DSRSC 263, and CDI 98 (see Fig. 3B and Supplementary Results S4.1).

Figure 3. Detection Rates and Distribution of Detection Conditions Across Depression Scales. Panel A shows the detection rates of depressive symptoms across four self-report scales. The x-axis represents the names of the scales, and the y-axis shows the corresponding detection rates. The different colors represent the detection rates for various age groups across the scales. Panel B illustrates the distribution of participants based on how many scales identified them as positive for depression. The depth of color reflects the number of scales on which participants screened positive. The darkest shade represents

participants who screened positive on all four scales. Medium-dark shades represent participants who screened positive on three scales. Medium-light shades represent participants who screened positive on two scales. The lightest shade represents participants who screened positive on only one scale. The area of each rectangle is proportional to the number of participants within each screening combination, and labels within each rectangle indicate the combination and corresponding sample size.

Exploratory Graph Analysis (EGA) revealed three communities among the 31 symptoms covered by these four scales (see Fig. 4 [Figure 4: see original paper]), providing empirical evidence for the heterogeneity among these symptoms (see Fig. 4; see Supplementary Results S4.2 for details).

Figure 4. The EGA results for the merged items. Different colors represent different communities. Community 1 was labeled as "MOOD," Community 2 as "COGNITION," and Community 3 as "FUNCTION." Sad, M1; Anhedonia, M2; Psychological worry, M3; Cry, M4; Psychic anxiety, M5; Fatigue, M6; Appetite change, M7; Somatic worry, M8; Feeling of loneliness, M9; Poor sleep, M10; Gastrointestinal, M11; Low mood, M12; Blue, M13; Feeling of hopelessness, C1; Psychological inferiority, C2; Self blame, C3; Felt people disliked me, C4; Running away from home, C5; Interest pleasure loss, C6; Feeling of worthlessness, C7; Negative body perception, C8; Like talking with family, C9; Reduced socialization, C10; Inferiority self confidence, C11; Bad all the time, C12; Learning difficulties, F1; Loss of energy, F2; Concentration, F3; Retardation and agitation, F4; Never have fun at school, F5; Indecisiveness, F6. Direct symptom-name plots can be found in Supplementary Results S4.2.

Discussion

Our study is the first systematic investigation of the heterogeneity of depression scales for children and adolescents in a non-Western cultural context. We identified 84 symptoms across 27 scales drawn from four recent meta-analyses [**?**, **?**, **?**, **?**, **?**] and found a weak overlap among them. This heterogeneity was further validated by re-analyzing a large dataset in which all participants completed four commonly used depression scales that cover 39 unique symptoms in total. We found that the four scales each identified different subsets of participants. Our analyses of the scale data revealed that these 39 symptoms formed more than one dimension. Importantly, we also found that only a few scales included youth- or cultural-specific symptoms. Given the importance of measurement tools in understanding, diagnosis, and intervention of depression [**?**, **?**] and the fact that depression is an increasing social burden for developing countries [**?**, **?**], understanding the heterogeneity of rating scales used in developing countries is of great importance for addressing mental health issues in these regions.

The pronounced heterogeneity observed across these 27 Chinese depression rating scales may be attributed to multiple reasons. Firstly, as many scales in-

cluded in our analyses were translated from Western scales, the two key sources of heterogeneity identified by Fried (2017) apply to our findings here. First, different psychopathological perspectives were different when these scales were developed. For instance, Beck' s Depression Inventory (BDI) emphasized cognitive symptoms in line with Beck' s theory of depression, whereas Hamilton' s Rating Scale (HAMD) highlighted somatic and anxiety-related symptoms that were easier to capture in clinician ratings. Second, different purposes of these scales. For example, the Center for Epidemiological Studies Depression Scale (CES-D) was designed for epidemiological screening in the general population, while the Hamilton Depression Rating Scale for Depression was originally intended to measure severity in clinically diagnosed patients. Moreover, because we included more scales, it is not surprising that heterogeneity was greater than that reported by Fried (2017).

A unique reason behind the heterogeneity found in our study is that some scales included age- and culture-specific items. For example, the Chinese College Student Mental Health Scale (CCSMHS) includes items such as "feeling of repression" and "want to take advantage," while the Adolescent Depression Inventory (ADI) includes symptoms such as "lack of patience" and "mood swings" . These items are usually idiosyncratic symptoms and therefore increase the heterogeneity of the rating scales. However, how to handle these items requires extreme care because including these items may increase the validity of these rating scales under their contexts. A growing body of cross-cultural and developmental research has shown that the subjective experience and symptoms of depression are shaped by both cultural factors and developmental stage [**?**, **?**, **?**, **?**]. For instance, symptoms such as "worry" or "thinking too much" are reported more frequently in South Asian, Southeast Asian, and Sub-Saharan African populations, suggesting culturally specific expressions of depression that may warrant inclusion in measurement instruments [**?**, **?**].

Beyond symptoms, cultural context shapes how depression is explained and acted upon. For instance, depression is viewed as a sign of personal weakness that does not require medical care in Australian First Nations \cite{Fusar-Poli et al., 2023}. Symptoms of depression may also differ at different ages. Previous data suggested that vegetative symptoms (e.g., appetite and weight changes, fatigue, insomnia) are more common in adolescents than adults, while anhedonia and concentration difficulties are more prominent in adults [**?**, **?**]. Taking together, the heterogeneity caused by items in scales developed by Chinese researchers is necessary given the importance of social context and age [**?**, **?**].

Our findings have important implications for researchers, practitioners, and policy-makers in non-Western contexts. For researchers, closer attention should be paid to the rating scales that are used in child and adolescent populations. First, it is important to identify core depressive symptoms that are most prevalent, impairing, and clinically relevant among the youth population in local settings. For example, in the large Sequenced Treatment Alternatives to Relieve Depression (STAR*D), network analysis identified sad mood, diminished

interest/pleasure, energy loss, and concentration problems as central symptoms, which are most strongly associated with functional impairment [?, ?]. However, the exact symptoms related to functional impairment might differ across age groups and cultures [?, ?, ?]. Second, developing methods to harmonize different rating scales to maximize existing datasets that used different rating scales. Recent work has demonstrated that a common metric can be developed through item response theory to equate multiple depression and anxiety scales in adolescent samples, providing scores with acceptable precision in the clinically relevant range [?, ?]. This approach allows researchers to pool data across cohorts and designs without being constrained to a single instrument. Third, researchers may leverage cross-cultural collaborations, especially big-team science [?, ?], to curate an open-access "measure hub" that includes existing rating scales, their translations and adaptations [?, ?]. Such a hub would facilitate adaptation of widely used scales while preserving comparability. Similarly, researchers in the field may consider following the Standardisation Of Behavior Research (SOBER) guidelines [?, ?, ?]—for instance, demonstrating non-redundancy when creating new measures, adhering to established protocols, and justifying any modifications with independent validity evidence—can help prevent further fragmentation and improve reproducibility. For practitioners, administering multiple depression scales in parallel and examining the average effect across all administered scales is recommended [?, ?]. Our empirical data have revealed substantial variation in detection rates across four scales, which suggests that relying on one scale might be risky. Thus, using multiple scales simultaneously can enhance the robustness of findings by enabling cross-validation across tools [?, ?]. Using multiple scales is already an established standard in certain fields, such as clinical trials, where one outcome is typically designated as primary and given analytic precedence over secondary instruments [?, ?].

For policy-makers, our findings suggest that the heterogeneity in rating scales is substantial and the large-scale survey' s findings that rely on single scales should be interpreted with care. Sun et al. (2025) have pointed out that the depression scales in one of the large-scale surveys in China ignored the pressure from school. Our results confirmed that while there exist scales developed by Chinese scholars, few studies used these scales in practice (see Fig. S2 for citations of 27 scales included in our content analysis). To promote standardization and reduce redundancy, policy-makers and professional societies could establish committee-set standards for youth depression assessment [?, ?]. Similar to how diagnostic criteria in the DSM and ICD are determined by expert committees, such standards could define core symptom domains and outline how they should be assessed in specific cultural contexts. Involving researchers, clinicians, educators, and young people themselves would ensure that these standards remain transparent, inclusive, and continuously updated to reflect developmental and cultural changes. Finally, our findings also call for funders to support the development, validation, maintenance, and dissemination of depression scales.

Several limitations of the current study should be noted. First, the content analysis in Study 1 is inherently subjective; it is possible that a different research

team might produce slightly different outcomes due to variations in interpretation or experience. In line with the principles of open science, we made all coding materials and analytic datasets publicly available to facilitate replication and further evaluation by other researchers. Second, all empirical data in Study 2 were drawn exclusively from one hospital in Beijing, China; thus, the generalizability of our findings to other regions and cultures needs to be examined by future studies.

Conclusion

This study systematically evaluated the heterogeneity of depression scales used for children and adolescents in China. Drawing on content analysis and empirical data, we identified substantial differences across scales in terms of symptom coverage, item formulation, and screening outcomes. This heterogeneity reflects not only inconsistencies across existing tools, but also cultural- and developmental specificities embedded in depression. We call for researchers, practitioners, and policy-makers in non-Western contexts to pay attention to the heterogeneity of depression scales and address this issue with care for better understanding, assessing, and intervening in depression.

References

Ali, G.-C., Ryan, G., & De Silva, M. J. (2016). Validated Screening Tools for Common Mental Disorders in Low and Middle Income Countries: A Systematic Review. *PLOS ONE*, 11(6), e0156939. https://doi.org/10.1371/journal.pone.0156939

Anvari, F., Alsalti, T., Oehler, L. A., Hussey, I., Elson, M., & Arslan, R. C. (2025). Defragmenting psychology. *Nature Human Behaviour*, 9(5), 836–839. https://doi.org/10.1038/s41562-025-02138-0

Cao, J. (2023). Promote the comprehensive development of students' physical and mental health. http://www.moe.gov.cn/jyb_{xwfb}/s271/202305/t20230511_{1059225}.html

Chan, D. W. (1997). Depressive symptoms and perceived competence among Chinese secondary school students in Hong Kong. *Journal of Youth and Adolescence*, 26(3), 303–319. https://doi.org/10.1007/s10964-005-0004-4

Chen, Y., Zhang, Y., & Yu, G. (2022). Prevalence of mental health problems among college students in mainland China from 2010 to 2020: A meta-analysis. *Advances in Psychological Science*, 30(5). https://doi.org/10.3724/SP.J.1042.2022.00991

Chen, Z., Hu, B., Liu, X., Becker, B., Eickhoff, S. B., Miao, K., Gu, X., Tang, Y., Dai, X., Li, C., Leonov, A., Xiao, Z., Feng, Z., Chen, J., & Chuan-Peng, H. (2023). Sampling inequalities affect generalization of neuroimaging-based diagnostic classifiers in psychiatry. *BMC Medicine*, 21(1). https://doi.org/10.1186/s12916-023-02941-4

Christensen, A. P., & Golino, H. (2021). Estimating the Stability of Psychological Dimensions via Bootstrap Exploratory Graph Analysis: A Monte Carlo

Simulation Tutorial. *Psych*, 3(3). https://doi.org/10.3390/psych3030032

Deng, H., Wen, F., Xu, H., Yang, H., Yan, J., Zheng, Y., Cui, Y., & Li, Y. (2023). Prevalence of affective disorders in Chinese school-attending children and adolescents aged 6–16 based on a national survey by MINI-Kid. *Journal of Affective Disorders*, 331, 192–199. https://doi.org/10.1016/j.jad.2023.03.060

Dreher, A., Hahn, E., Diefenbacher, A., Nguyen, M. H., Böge, K., Burian, H., Dettling, M., Burian, R., & Ta, T. M. T. (2017). Cultural differences in symptom representation for depression and somatization measured by the PHQ between Vietnamese and German psychiatric outpatients. *Journal of Psychosomatic Research*, 102, 71–77. https://doi.org/10.1016/j.jpsychores.2017.09.010

Elson, M., Hussey, I., Alsalti, T., & Arslan, R. C. (2023). Psychological measures aren' t toothbrushes. *Communications Psychology*, 1(1). https://doi.org/10.1038/s44271-023-00026-9

Evans, L., Haeberlein, K., Chang, A., & Handal, P. (2021). Convergent Validity and Preliminary Cut-Off Scores for the Anxiety and Depression Subscales of the DASS-21 in US Adolescents. *Child Psychiatry & Human Development*, 52.

Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A., Dutra, N., Basnight-Brown, D., & IJzerman, H. (2023). The benefits, barriers, and risks of big-team science. *Perspectives on Psychological Science*, 18(3), 607–623.

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, 208, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, 189, 314–320. https://doi.org/10.1016/j.jad.2015.09.005

Fried, E. I., Flake, J. K., & Robinaugh, D. J. (2022). Revisiting the theoretical and methodological foundations of depression measurement. *Nature Reviews Psychology*, 1(6), 358–368. https://doi.org/10.1038/s44159-022-00050-2

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don' t add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, 13(1), 72. https://doi.org/10.1186/s12916-015-0325-4

Fusar-Poli, P., Estradé, A., Stanghellini, G., Esposito, C. M., Rosfort, R., Mancini, M., Norman, P., Cullen, J., Adesina, M., Jimenez, G. B., da Cunha Lewin, C., Drah, E. A., Julien, M., Lamba, M., Mutura, E. M., Prawira, B., Sugianto, A., Teressa, J., White, L. A., ···Maj, M. (2023). The lived experience of depression: A bottom-up review co-written by experts by experience and academics. *World Psychiatry*, 22(3), 352–365. https://doi.org/10.1002/wps.21111

Golino, H. F., & Christensen, A. P. (2024). EGAnet: Exploratory Graph Analysis –A framework for estimating the number of dimensions in multivariate data

using network psychometrics. https://doi.org/10.32614/CRAN.package.EGAnet

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, 12(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H. F., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, 25(3), 292.

Gong, X., Xie, X., Xu, R., & Luo, Y. (2010). Psychometric Properties of the Chinese Versions of DASS-21 in Chinese College Students. *Chinese Journal of Clinical Psychology*, 18(4). https://doi.org/10.16128/j.cnki.1005-3611.2010.04.020

Haroz, E. E., Ritchey, M., Bass, J. K., Kohrt, B. A., Augustinavicius, J., Michalopoulos, L., Burkey, M. D., & Bolton, P. (2017). How is depression experienced around the world? A systematic review of qualitative literature. *Social Science & Medicine*, 183, 151–162. https://doi.org/10.1016/j.socscimed.2016.12.030

Hayes, J., Carvajal-Velez, L., Hijazi, Z., Ahs, J. W., Doraiswamy, P. M., El Azzouzi, F. A., Fox, C., Herrman, H., Gornitzka, C. P., Staglin, B., & Wolpert, M. (2023). You Can' t Manage What You Do Not Measure—Why Adolescent Mental Health Monitoring Matters. *Journal of Adolescent Health*, 72(1), S7–S8. https://doi.org/10.1016/j.jadohealth.2021.04.024

Huang, X., Zhang, Y., & Yu, G. (2022). Prevalence of mental health problems among primary school students in Chinese mainland from 2010 to 2010: A meta-analysis. *Advances in Psychological Science*, 30(5). https://doi.org/10.3724/SP.J.1042.2022.00953

Krause, K. R., Chung, S., Sousa Fialho, M. da L., Szatmari, P., & Wolpert, M. (2021). The challenge of ensuring affordability, sustainability, consistency, and adaptability in the common metrics agenda. *The Lancet Psychiatry*, 8(12), 1094–1102. https://doi.org/10.1016/S2215-0366(21)00122-X

Li, F., Cui, Y., Li, Y., Guo, L., Ke, X., Liu, J., Luo, X., Zheng, Y., & Leckman, J. F. (2022). Prevalence of mental disorders in school children and adolescents in China: Diagnostic data from detailed clinical assessments of 17,524 individuals. *Journal of Child Psychology and Psychiatry*, 63(1). https://doi.org/10.1111/jcpp.13445

Lu, B., Lin, L., & Su, X. (2024). Global burden of depression or depressive symptoms in children and adolescents: A systematic review and meta-analysis. *Journal of Affective Disorders*, 354, 553–562. https://doi.org/10.1016/j.jad.2024.03.074

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales

of Cultural and Psychological Distance. *Psychological Science*, 31(6), 678–701. https://doi.org/10.1177/0956797620916782

Nagata, J. M., Hathi, S., Ferguson, B. J., Hindin, M. J., Yoshida, S., & Ross, D. A. (2018). Research priorities for adolescent health in low- and middle-income countries: A mixed-methods synthesis of two separate exercises. *Journal of Global Health*, 8(1). https://doi.org/10.7189/jogh.08.010501

Nili, A., Tate, M., Barros, A., & Johnstone, D. (2020). An approach for selecting and using a method of inter-coder reliability in information management research. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2020.102154

Nurismawan, Ach. S., Lestary, Y. D., Purwoko, B., Alfaried, H. Z., & Nafisah, K. (2024). Unraveling the dangers of mental health self-diagnosis: A study on the phenomenon of adolescent self-diagnosis in junior high schools. *Jurnal Bimbingan Dan Konseling*, 11(1), 31–38.

Park, S., Jang, E. Y., Xiang, Y., Kanba, S., Kato, T. A., Chong, M., Lin, S., Yang, S., Avasthi, A., Grover, S., Kallivayalil, R. A., Udomratn, P., Chee, K. Y., Tanra, A. J., Tan, C., Sim, K., Sartorius, N., Park, Y. C., & Shinfuku, N. (2020). Network analysis of the depressive symptom profiles in Asian patients with depressive disorders: Findings from the Research on Asian Psychotropic Prescription Patterns for Antidepressants (REAP-AD). *Psychiatry and Clinical Neurosciences*, 74(6), 344–353. https://doi.org/10.1111/pcn.12989

Rice, F., Riglin, L., Lomax, T., Souter, E., Potter, R., Smith, D. J., Thapar, A. K., & Thapar, A. (2019). Adolescent and adult differences in major depression symptom profiles. *Journal of Affective Disorders*, 243, 175–181. https://doi.org/10.1016/j.jad.2018.09.015

Sakuma, K.-L. K., Sun, P., Unger, J. B., & Johnson, C. A. (2010). Evaluating Depressive Symptom Interactions on Adolescent Smoking Prevention Program Mediators: A Mediated Moderation Analysis. *Nicotine & Tobacco Research*, 12(11), 1099–1107. https://doi.org/10.1093/ntr/ntq156

Schlechter, P., Ford, T. J., & Neufeld, S. A. S. (2023). The development of depressive symptoms in older adults from a network perspective in the English Longitudinal Study of Ageing. *Translational Psychiatry*, 13(1), 363. https://doi.org/10.1038/s41398-023-02659-0

Seppänen, M., Lankila, T., Auvinen, J., Miettunen, J., Korpelainen, R., & Timonen, M. (2022). Cross-cultural comparison of depressive symptoms on the Beck Depression Inventory-II, across six population samples. *BJPsych Open*, 8(2), e46. https://doi.org/10.1192/bjo.2022.13

Shorey, S., Ng, E. D., & Wong, C. H. J. (2022). Global prevalence of depression and elevated depressive symptoms among adolescents: A systematic review and meta-analysis. *British Journal of Clinical Psychology*, 61(2), 287–305. https://doi.org/10.1111/bjc.12333

Stallwood, E., Monsour, A., Rodrigues, C., Monga, S., Terwee, C., Offringa, M., & Butcher, N. J. (2021). Systematic review: The measurement properties of the children' s depression rating scale—revised in adolescents with major depressive disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 60(1), 119–133. https://doi.org/10.1016/j.jaac.2020.10.009

Stockings, E., Degenhardt, L., Lee, Y. Y., Mihalopoulos, C., Liu, A., Hobbs, M., & Patton, G. (2015). Symptom screening scales for detecting major depressive disorder in children and adolescents: A systematic review and meta-analysis of reliability, validity and diagnostic utility. *Journal of Affective Disorders*, 174, 447–463. https://doi.org/10.1016/j.jad.2014.11.061

Su, L., Wang, K., Zhu, Y., Luo, X., & Yang, Z. (2003). Norm of The Depression Self-rating Scale for Children in Chinese Urban Children. *Chinese Mental Health Journal*, 17(8), 547–549.

Sun, J., Rose-Clarke, K., Bao, Y., Wang, Z., & Lu, L. (2025). Child and adolescent mental health policy advancement in China. *The Lancet Psychiatry*. https://doi.org/10.1016/S2215-0366(25)00240-8

Sunderland, M., Olsen, N., Visontay, R., Chapman, C., Mewton, L., Stapinski, L., Newton, N., Teesson, M., & Slade, T. (2024). "One Metric to Rule Them All" : A Common Metric for Symptoms of Depression and Generalized Anxiety in Adolescent Samples. *Clinical Psychological Science*, 12(1), 147–160. https://doi.org/10.1177/21677026231168564

Veal, C., Tomlinson, A., Cipriani, A., Bulteau, S., Henry, C., Müh, C., Touboul, S., De Waal, N., Levy-Soussan, H., Furukawa, T. A., Fried, E. I., Tran, V.-T., & Chevance, A. (2024). Heterogeneity of outcome measures in depression trials and the relevance of the content of outcome measures to patients: A systematic review. *The Lancet Psychiatry*, 11(4), 285–294. https://doi.org/10.1016/S2215-0366(23)00438-8

Wahid, S. S., Ottman, K., Bohara, J., Neupane, V., Fisher, H. L., Kieling, C., Mondelli, V., Gautam, K., & Kohrt, B. A. (2022). Adolescent perspectives on depression as a disease of loneliness: A qualitative study with youth and other stakeholders in urban Nepal. *Child and Adolescent Psychiatry and Mental Health*, 16(1). https://doi.org/10.1186/s13034-022-00481-y

Yu, D., & Li, X. (2000). Preliminary Use of the Children' s Depression Inventory in China. *Chinese Mental Health Journal*, 14(4), 225–227.

Yu, X., Zhang, Y., & Yu, G. (2022). Prevalence of mental health problems among senior high school students in mainland of China from 2010 to 2020: A meta-analysis. *Advances in Psychological Science*, 30(5). https://doi.org/10.3724/SP.J.1042.2022.00978

Zhang, M., & He, Y. (2015). *Handbook of Rating Scales in Psychiatry*. Hunan Science and Technology Press.

ChinaRxiv 中X

Zhang, Y., Jin, J., & Yu, G. (2022). Prevalence of mental health problems among junior high school students in Chinese mainland from 2010 to 2020: A meta-analysis. *Advances in Psychological Science*, 30(5). https://doi.org/10.3724/SP.J.1042.2022.00965

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.