

BMJ Open Development and validation of brief scales to measure emotional and behavioural problems among Chinese adolescents

Minxue Shen,^{1,2} Ming Hu,² Zhenqiu Sun²

To cite: Shen M, Hu M, Sun Z. Development and validation of brief scales to measure emotional and behavioural problems among Chinese adolescents. *BMJ Open* 2017;7:e012961. doi:10.1136/bmjopen-2016-012961

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2016-012961>).

Received 10 June 2016
Revised 19 September 2016
Accepted 5 October 2016



CrossMark

¹Xiangya Hospital, Central South University, Changsha, China

²Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

Correspondence to

Ming Hu; xysm2011@yeah.net

ABSTRACT

Objectives: To develop and validate brief scales to measure common emotional and behavioural problems among adolescents in the examination-oriented education system and collectivistic culture of China.

Setting: Middle schools in Hunan province.

Participants: 5442 middle school students aged 11–19 years were sampled. 4727 valid questionnaires were collected and used for validation of the scales. The final sample included 2408 boys and 2319 girls.

Primary and secondary outcome measures: The tools were assessed by the item response theory, classical test theory (reliability and construct validity) and differential item functioning.

Results: Four scales to measure anxiety, depression, study problem and sociality problem were established. Exploratory factor analysis showed that each scale had two solutions. Confirmatory factor analysis showed acceptable to good model fit for each scale. Internal consistency and test–retest reliability of all scales were above 0.7. Item response theory showed that all items had acceptable discrimination parameters and most items had appropriate difficulty parameters. 10 items demonstrated differential item functioning with respect to gender.

Conclusions: Four brief scales were developed and validated among adolescents in middle schools of China. The scales have good psychometric properties with minor differential item functioning. They can be used in middle school settings, and will help school officials to assess the students' emotional/behavioural problems.

INTRODUCTION

Adolescence is a time of profound biological and social transition during which new behaviours are developed that can either benefit the health and social adaptation of youth or, alternatively, undermine adjustment in adulthood.¹ Adolescents have to cope with increasing independence and the growing importance of social relationships, while developing and exercising self-control. However,

Strengths and limitations of this study

- First study to develop and validate brief tools for assessing common emotional and behavioural problems among Chinese adolescents.
- Modern test theory and classical test theory are used to validate the tool.
- The scales are specifically relevant to the examination-oriented education system and collectivistic culture of China.
- A convenient sampling method was used.
- Diagnoses were not determined by psychiatrists.

difficulties in the development of executive control in adolescence can lead to a lack of balance in the regulation of cognition, emotions and behaviours when dealing with negative thoughts and feelings, and then result in anxiety and depression.² This is an important issue, especially in low-income and middle-income countries where enormous environmental threats are more common (eg, poverty, war, internal conflicts, sex trafficking, early pregnancy and marriage, absence of access to education).

Adolescents, defined by the WHO as those between 10 and 19 years of age, represent an estimated 1.2 billion of the world's population.³ A meta-analysis that synthesised data from over 60 000 adolescents aged 13–18 years estimated the prevalence of depression to be 6% in the community.⁴ Anxiety disorders have an estimated cumulative prevalence of 32% among adolescents aged 13–18 years according to numerous population-based studies.⁵ Mental disorders also contribute to heavy social burdens. Globally, neuropsychiatric disorders accounts for 45% of years lost due to disability for adolescents.⁶ Up to 20% of young adults have a disabling mental illness, and 50% of adult mental health disorders experience their onset in adolescence.⁷ Study and sociality problems are associated

with anxiety and depression disorders in adolescents, while these disorders per se lead to negative outcomes including behavioural problems, poor school performance and impaired social and family functioning.^{8,9}

Culture influences the sources of distress, the form of illness experience, symptomatology, the interpretation of symptoms, modes of coping with distress, help-seeking and the social response to distress and disability. In China, academic pressure under the examination-oriented education system, cultural differences between China and Western societies with respect to the view of social and sexual relationships, as well as limited mental health literacy and social inequity experienced by massive younger generations of internal migrants resulted in increased prevalence of anxiety, depression and other mental disorders.^{10–12} The examination-oriented education system of China results in a score-based friend-making criterion: adolescents prefer to make friends with those who perform better in examinations, and ignore those with poor performance. This further aggravates the association of anxiety and depression with study and sociality problems among them. However, since individual socioemotional well-being has traditionally been neglected in the collectivistic culture of China, adolescent mental and behavioural problems are not well identified, and have not received adequate attention from professionals and the public.¹³

Collectivism, as always being enhanced by Chinese parents and teachers, also results in differences in the expression of anxiety and depression. A collectivistic culture values harmony within the group, and the individual gain is considered to be less important than improvement of the social group.¹⁴ Embarrassment may be more common in collectivistic cultures because it is induced by external sanctions.¹⁵ ‘Taijin kyofusho’ (the fear of offending or embarrassing the other person) is an example of a culturally specific expression of anxiety in Asian countries.¹⁶ Biological evidence also showed that people who live in collectivist cultures are more likely than those in individualistic cultures to have a form of the serotonin transporter gene that correlates with higher rates of anxiety and depression.¹⁷

In addition, there are cultural differences with respect to treatment response. Stigmatisation of people with mental illness is especially pronounced in China.¹⁸ Family members try to conceal any history of mental illness within the family to avoid any negative impact on the family and potential of the young person to get married. A study examined culture-related influences on willingness to seek treatment for anxiety in first-generation and second-generation students of Chinese heritage and their European-heritage counterparts, and found that first-generation Chinese participants were significantly less willing to seek treatment.¹⁹ The reluctance was associated with greater Chinese-heritage acculturation rather than perceiving symptoms as less impairing.

A psychological scale seeks to identify and evaluate patients who may have current disorders but have not

sought treatment.²⁰ Currently, widely used mental/behaviour problem scales for children and adolescents include the Achenbach Child Behavior Checklist,²¹ Personality Diagnostic Questionnaire,²² Rutter’s Behavior Scale,²³ Spence Children’s Anxiety Scale,²⁴ Zung’s Self-Rating Anxiety Rating Scale (SAS),²⁵ Zung’s Self-Rating Depression Scale (SDS),²⁶ Children’s Depression Inventory,²⁷ Child and Adolescent Psychiatric Assessment,²⁸ Hospital Depression and Anxiety Scale,²⁹ etc. In China, however, most studies employed translated versions of foreign scales,^{30–34} and no dedicated scale has been developed, validated and used to screen the common emotional and behaviour problems among adolescents in middle school. Owing to the cross-cultural differences with respect to the formation and expression of mental problems as mentioned above, in the current study, we developed and validated four brief scales according to the characteristics of adolescents in the examination-oriented education system and collectivistic culture of China.

METHODS

Study design

The study had a cross-sectional design. The participants were junior school (grade 7th–9th) and high school (grade 10th–12th) students aged 11–19. A stratified cluster sampling frame was used. Two cities (Changsha and Shaoyang) and two counties (Liuyang and Ningxiang) were selected from Hunan province through a convenience sampling method. In each city or county, two junior schools and two high schools were selected using a random number table. In each school, two classes were selected from each grade using a random number table. All students in selected classes were recruited through a questionnaire survey.

Scale development

Based on literature review and expert advice, common emotional and behavioural problems among Chinese middle school students were selected (emotional problems, learning problems and interpersonal problems), and four scales were formed (anxiety, depression, study problems and sociality problem). Scale development was performed in four phases:

Phase I: Programmed decision processing was used to develop the scales by a nominal group. A total of 90 items was drafted by interviewing the nominal group.

Phase II: Individual questions were edited and redundant questions were eliminated by a focus group consisting of 10 experts of child and adolescent psychology (n=3), epidemiology (n=2), biostatistics (n=3) and medical sociology (n=2). An initial pool of 59 items was derived. The responses to all items were graded on a four-point scale (1=never/occasionally; 2=a little time; 3=quite some time; 4=most of the time). Responses were transformed into raw scores, which signified the severity of a problem.

Phase III: A pilot study was conducted in 538 middle school students (one class per grade from a junior school and a high school in Changsha and Ningxiang, respectively). The items were selected by statistical methods as follows: (1) t-test. Participants were ranked by the score on the scale, and a high-score group and a low-score group were derived, respectively, according to percentiles (P75 and P25). The score of each item was compared using Student's t-test. Any item with no statistical difference between the two groups was eliminated. (2) Correlation coefficient. Any item with a Pearson's correlation coefficient <0.40 with the scale score was eliminated. After the pilot test, four scales consisting of 46 items were derived. (3) Factor analysis. Any item with a factor loading <0.40 was eliminated. (4) Any item with three options that presented a selection rate <10% was eliminated. Details of the pilot study can be found in our published paper.³⁵

Phase IV: The scales were tested among 5442 middle school students. A total of 220 students received the test again 1-week after the initial test. The construct validity, reliability, item performance and differential item functioning (DIF) were assessed.

Statistical analysis

The Kaiser-Meyer-Olkin (KMO) test was used to evaluate the adequacy of exploratory factor analysis (EFA), and a KMO value >0.8 indicates that factor analysis will be useful. EFA was used to examine the number of factors in each scale. Factors with an eigenvalue >1.0 were selected. The quartimax rotation was used to achieve rotated factor loadings.

Under the a priori hypothesis that each scale measures a distinct latent trait, construct validity was assessed by confirmatory factor analysis (CFA) for each scale respectively, using structured equation models. For each scale, the number of dimensions was determined by the EFA result. Goodness-of-fit of CFA was assessed by the root mean square error of approximation (RMSEA), comparative fit index (CFI) and Tucker-Lewis index (TLI).

Values of the CFI and TLI ≥ 0.90 represent an acceptable fit, and >0.95 represents a good fit. Values of the RMSEA ≤ 0.08 represent an acceptable fit, and <0.05 represents a good fit.³⁶ Factor loadings were reported.

Concurrent validity of scales was assessed by Zung's Self-Rating Anxiety Scale (SAS), Zung's Self-Rating Depression Scale (SDS), and the interpersonal sensitivity dimension of Chinese SCL-90-R, respectively, using Pearson's correlation coefficients. Concurrent validity of the study problem scale was not evaluated owing to the lack of relevant scales.

Reliability of scales was assessed by Cronbach's α coefficient, Spearman-Brown's split-half coefficient and test-retest reliability. Item performance was assessed by the two-parameter polytomous item response theory (IRT) models. IRT is a family of associated mathematical models that relate latent traits (ability) to the probability of responses to items in an assessment, and it has been

widely used in health assessment.^{37 38} It describes a non-linear relationship between binary, ordinal or categorical responses and the latent trait (mental/behavioural problems in this study). Equation (1) specifies a polytomous IRT model, which is used for items with multiple categories (eg, Likert-type).

$$P(Y_{ik} = c | \theta_i, b_k) = \frac{\exp[a_k(\theta_i - b_{k,c-1})]}{1 + \exp[a_k(\theta_i - b_{k,c-1})]} - \frac{\exp[a_k(\theta_i - b_{k,c})]}{1 + \exp[a_k(\theta_i - b_{k,c})]} \quad (1)$$

In this model, the probability (P) of scoring in a specific category (c) is modelled by the probability of responding in this category minus the probability of responding in the next category. $b_{k,c}$ is the upper grade difficulty parameter (the point on the ability scale that corresponds to a probability of a certain response of 0.5) for category c of item k, and a_k is the discrimination parameter (estimates how well an item can differentiate among respondents with different levels of ability) for item k. Acceptable a_k should be above 0.5, and appropriate mean b_k should be between -3.0 and 3.0.³⁹

Test information function (TIF) describes the precision of the measure. A measure has most discriminative power among participants with ability that corresponds to the peak of the TIF curve.

Measurement invariance of the scales between genders was evaluated using the DIF test. Gender differences of discrimination (Δa_i) and difficulty (Δb_i) parameters for all items were examined using the χ^2 tests with a significant level of 0.001.

IRT parameters were estimated using the Bock-Aitkin procedure⁴⁰ in IRTPRO 3 (Scientific Software International, Lincolnwood). Other analyses were performed using SAS 9.4 (SAS Institute, Cary, North Carolina, USA). The significance level for DIF was 0.001 owing to the large sample size and Bonferroni correction for multiple comparisons; for other statistical tests, the α was 0.01.

Ethics statement

The methods were carried out in accordance with the STROBE statement.⁴¹ The purpose and implication of the survey were explained by the investigators. Written informed consents were obtained from all parents or main caregivers of the students.

RESULTS

In all, 5442 middle school students were sampled, 5273 (97%) returned the questionnaires and 4727 (87%) completed the survey without apparent logical errors and missing values on items. The average age of the participants was 14.9 ± 1.9 and ranged from 11 to 19 years. The demographic characteristics of the students and their parents are shown in table 1.

Table 1 Demographic characteristics of participants

Characteristics	N (%)
City/county	
Changsha	1066 (22.6)
Shaoyang	1268 (26.8)
Liuyang	1226 (25.9)
Ningxiang	1167 (24.7)
Gender	
Male	2408 (50.9)
Female	2319 (49.1)
Age (years)	
11–14	2057 (43.5)
15–19	2670 (56.5)
Grade	
7th–9th	2459 (52.0)
10th–12th	2268 (48.0)
Ethnic group	
Han	4649 (98.3)
Minority	78 (1.7)
Parent's marriage status	
Married	3873 (81.9)
Divorced/separated/widowed	212 (4.5)
Declined/missing	642 (13.6)
Father's education level	
Primary school	286 (6.1)
Junior School	1445 (30.6)
High School	1186 (25.1)
College and above	1123 (23.7)
Declined/missing	687 (14.5)
Mother's education level	
Primary school	414 (8.8)
Junior School	1673 (35.4)
High School	1194 (25.3)
College and above	776 (16.4)
Declined/missing	670 (14.2)

The KMO test values for four scales were 0.893 (anxiety), 0.864 (depression), 0.870 (study problem) and 0.870 (sociality problem), respectively, all with Bartlett's sphericity test $p < 0.01$. The results of EFA are shown in [table 2](#). For each scale, two factors with an eigenvalue > 1.0 were identified. For the anxiety scale, the two factors signified general symptoms and sleep-related symptoms, respectively. For the depression scale, the two factors loaded on different symptoms that were psychologically difficult to group. For the study problem, the two factors referred to interest of study and study/exams stress, respectively. For the sociality problem, the two factors signified sociality problems at home and at school, respectively.

The results of CFA are shown in [table 3](#). For each scale, two dimensions were identified by EFA, and were used for CFA. Mixed evidence was found for the dimensionality tested for each of the scales, that is, there was acceptable to good RMSEA and CFI for all scales but unacceptable TLI (< 0.9) for the sociality scale. Despite the good model fit, local dependence was identified between the item A6 and A7, and between D9 and D11.

Table 2 Exploratory factor analyses for the four scales

Scale	Factors	Per cent of variance	Items
Anxiety	1. General symptoms	36.5	A1–A9
	2. Sleeping status	9.5	A10, A11
Depression	1. General symptoms	29.5	D1–D4, D6, D8, D10, D11
	2. Other symptoms	9.6	D5, D7, D9, D12
Study	1. Interest of study	36.7	ST1–ST6, ST11–ST13
	2. Stress coping	10.3	ST7–ST10
Sociality	1. At home	41.4	SO1–SO4
	2. At school	15.7	SO5–SO10

Table 3 Construct validity and reliability of the four scales

	Anxiety	Depression	Study	Sociality
CFA model fit				
RMSEA	0.054	0.032	0.062	0.079
CFI	0.953	0.974	0.934	0.902
TLI	0.938	0.968	0.919	0.861
Reliability				
Cronbach's α	0.804	0.759	0.846	0.735
Split-half reliability	0.717	0.750	0.808	0.491
Test–retest reliability *	0.725	0.764	0.794	0.725

*Test–retest reliability was tested among 220 students. CFA, confirmatory factor analysis; CFI, comparative fit index; RMSEA, root mean square error of approximation; TLI, Tucker-Lewis index.

CFA models were modified accordingly, and the goodness of fit was improved slightly. Pearson's correlation coefficient and standardised regression weight (factor loading) of each item are shown in [table 4](#).

With concurrent validity, the anxiety scale had a correlation coefficient of 0.78 ($p < 0.0001$) with SAS, the depression scale had a coefficient of 0.79 with SDS ($p < 0.0001$), and the sociality problem scale had a coefficient of 0.47 ($p < 0.0001$) with an interpersonal sensitivity subscale of Chinese SCL-90-R.

Cronbach's α , Spearman-Brown's split-half coefficient and test–retest reliability of each scale are shown in [table 3](#). Overall, all scales showed good internal consistency and test–retest reliability (above 0.7).

The discrimination (a_i) and difficulty parameters (b_i) of the polytomous IRT models are presented in [table 4](#). All items exhibited an acceptable discrimination parameter ($a_i > 0.5$) and many had high discrimination power

Table 4 Psychometric parameters of items in the original scale

Scales/items	Pearson's correlation	Factor loading	IRT parameters					DIF test for gender	
			a _i	b ₁	b ₂	b ₃	b _i (mean)	Δa _i	Δb _i
Anxiety									
A1. I feel more nervous and anxious than usual.	0.61	0.56	1.40	-0.34	1.58	2.91	1.38	0.09	-0.09
A2. I feel afraid for no reason at all.	0.65	0.62	1.68	0.31	1.68	2.72	1.57	0.41 ⁺	-0.41 ⁺
A3. I get upset easily.	0.72	0.72	2.20	-0.82	0.70	1.89	0.59	0.12	-0.07
A4. I feel that I am going to be crazy.	0.70	0.70	2.28	0.45	1.46	2.21	1.38	0.28	-0.04
A5. I feel weak and get tired easily.	0.70	0.68	1.90	-0.44	0.63	2.10	0.76	0.20	-0.03
A6. I always feel my heart beating fast.	0.63	0.57	1.55	0.23	1.94	2.96	1.71	-0.13	0.39 [†]
A7. I always feel trouble with breathing.	0.62	0.56	1.66	0.93	2.22	3.18	2.11	-0.03	-0.11
A8. I am bothered by headaches, neck/back pain, or abdominal pain.	0.63	0.56	1.29	-0.27	1.43	2.81	1.32	-0.09	-0.13
A9. I have to empty my bladder often.	0.39	0.28	0.61	-0.38	3.25	6.25	3.04	-0.14	1.00
A10. I fall asleep easily and get a good night's rest.	0.40	0.32	0.56	-1.77	0.61	3.57	0.80	0.03	0.10
A11. I often have nightmares.	0.55	0.59	2.27	0.50	2.11	2.81	1.81	-0.34	0.08
Depression									
D1. I feel downhearted and blue.	0.64	0.67	1.89	-0.43	1.40	2.34	1.11	0.18	0.09
D2. I often feel like crying.	0.59	0.59	1.50	-0.21	1.23	2.33	1.12	0.37 ⁺	-0.62 ⁺
D3. I have trouble sleeping at night.	0.52	0.47	1.02	-0.13	1.66	3.00	1.51	0.07	0.15
D4. My heart beats faster than usual.	0.51	0.50	1.26	0.63	2.63	3.67	2.31	-0.08	0.45 ⁺
D5. I eat as much as I used to.	0.37	0.39	1.09	-0.99	0.52	2.16	0.56	0.09	0.12
D6. I get tired for no reason.	0.62	0.64	1.72	-0.32	1.20	2.22	1.03	0.23	-0.05
D7. My mind is as clear as it used to be.	0.50	0.66	1.55	-0.91	0.52	2.01	0.54	0.15	0.05
D8. I am restless and cannot keep still.	0.64	0.69	2.12	-0.33	1.22	2.21	1.03	0.02	0.13
D9. I feel hopeful about the future.	0.46	0.49	1.06	-0.20	1.42	3.23	1.48	0.07	-0.08
D10. I am more irritable than usual.	0.54	0.52	1.26	-0.56	1.37	2.55	1.12	0.07	0.10
D11. I feel I am inferior to my classmates.	0.56	0.48	1.07	-0.71	1.53	2.91	1.24	0.08	-0.24
D12. I am still interested in the things that I used to do.	0.39	0.42	0.99	-0.57	1.16	3.71	1.43	0.14	-0.20
Study problem									
ST1. I am not interested in studying.	0.62	0.61	1.89	-0.58	1.37	2.41	1.07	0.46 ⁺	-0.08
ST2. I have to compel myself to study.	0.60	0.57	1.45	-0.88	1.01	2.48	0.87	0.62 ⁺	-0.07
ST3. I am not good at arranging time for study.	0.65	0.64	1.58	-1.25	0.57	2.04	0.45	0.02	-0.08
ST4. I cannot concentrate in class.	0.68	0.69	2.01	-0.89	0.95	2.13	0.73	0.09	0.05
ST5. I never discuss questions with teachers or classmates.	0.51	0.46	0.96	-0.74	1.41	2.90	1.19	-0.11	0.45 ⁺
ST6. I hardly teach myself even if the teacher asks me to.	0.60	0.56	1.42	-0.80	0.92	2.11	0.74	0.06	0.08
ST7. I feel stressful when learning.	0.59	0.59	1.39	-1.50	0.38	1.62	0.17	0.13	0.12
ST8. I feel that I cannot achieve my parents' or teachers' expectation.	0.70	0.78	2.98	-0.95	0.25	1.09	0.13	-0.08	0.03
ST9. I become very nervous before examinations.	0.37	0.32	0.79	-0.73	1.71	3.25	1.41	0.25	-0.23
ST10. I have no confidence in accomplishing my schoolwork.	0.69	0.76	2.20	-0.54	0.78	1.60	0.62	0.29	-0.13
ST11. I am always absent from school.	0.27	0.26	1.43	2.77	4.14	4.63	3.85	0.17	0.29 ⁺
ST12. I feel that I cannot adapt myself to the current stage of learning.	0.69	0.69	1.80	-0.32	1.59	2.47	1.25	0.03	-0.17
ST13. I hate studying.	0.69	0.70	2.38	-0.10	1.58	2.26	1.25	0.07	0.01
Sociality problem									
SO1. I get along well with my parents.	0.66	0.77	2.87	-0.17	0.93	2.12	0.96	0.27	-0.07
SO2. I would like to talk to my parents when I have troubles.	0.63	0.64	1.61	-1.39	-0.57	0.80	-0.39	0.30	0.18
SO3. I feel that my parents are interfering in my life too much.	0.41	0.37	0.94	-0.60	1.85	2.98	1.41	-0.05	0.47 ⁺
SO4. I feel happy at home.	0.63	0.71	2.16	-0.51	0.54	1.85	0.63	-0.21	-0.06
SO5. I would like to talk to my teachers when I have troubles.	0.50	0.43	0.50	-5.52	-3.66	-0.14	-3.11	-0.15	-0.85

Continued

Table 4 Continued

Scales/items	Pearson's correlation	Factor loading	IRT parameters			DIF test for gender			
			a_i	b_1	b_2	b_3	b_i (mean)	Δa_i Δb_i	
SO6. I get along well with my classmates.	0.59	0.69	3.09	-0.09	1.25	2.48	1.21	0.38	-0.06
SO7. I feel that I can deal with the troubles among classmates well.	0.56	0.60	2.02	-0.50	0.69	2.21	0.80	0.18	-0.05
SO8. I feel that I am often misunderstood by my classmates.	0.36	0.25	0.59	-0.54	3.03	5.25	2.58	0.14	-0.40
SO9. I have friends that I can share my happiness and secrets with.	0.52	0.49	1.19	-0.26	0.96	2.17	0.96	0.13	0.40
SO10. I feel happy at school.	0.57	0.58	1.19	-1.12	0.51	2.61	0.67	-0.09	-0.09

a_i : discrimination parameter for item i . b_1 to b_3 : difficulty parameters of each category. b_i : mean difficulty parameter for categories 1–4 of item i . $\Delta a_i = a_i$ (girls) – a_i (boys), $\Delta b_i = b_i$ (girls) – b_i (boys).
* $p < 0.001$ (the significant level for DIF tests).
DIF, differential item functioning.

Most items had an appropriate difficulty parameter ($-3.0 < b_i < 3.0$) except A9, ST11 and SO5. As shown in figure 1, for the anxiety, depression and study scales, TIF reached a peak where students' ability was around 2.0; this indicates that the measurement was most discriminative among students with a high level of problems. The TIF of the sociality scale reached a peak among students with a moderate level of problem (ability around 0).

The DIF test indicated that 2/11, 2/12/, 4/13 and 2/10 items in anxiety, depression, study and sociality scales, respectively, showed significant ($p < 0.001$) DIF with respect to gender (table 4). A negative Δb_i indicated that girls were more likely to endorse a higher score than boys; and vice versa, a positive Δb_i indicated that boys were more likely to endorse a higher score. A mixed pattern of DIF was observed.

DISCUSSION

We developed and validated four short scales to measure emotional and behavioural problems among adolescents in middle schools of Hunan, China. In this study, we (1) developed the initial item pool based on a focus group, broad literature review and existing scales; (2) examined the psychometric properties of the scales in a pilot study among 538 middle school students; (3) examined the psychometric properties of the scales according to classical and modern test theories among a large sample of adolescents. This is the first systematic study to develop and validate short scales to measure emotional (anxiety and depression) and behavioural (study and sociality) problems among adolescents in China.

Cultural differences profoundly impact the sources of distress (study maladaptation in our case), formation and expression of anxiety/depression (characterised by amplified somatic symptoms among the Asian population) and sociality problems (collectivism in Chinese culture vs individualism in western cultures). As a result, the scales are specific in several aspects: (1) In the anxiety and depression scale, some somatic symptom items were not included in our scale to minimise the effect of amplification of somatic symptoms, because somatisation among Asians has been well documented in comparison studies.^{42–44} Sex-related items were not considered as well because of the inapplicability. Important and frequently reported items were retained, such as cardiopulmonary and vestibular symptoms, pain and fatigue. (2) The study scale measured the interest and stress related to the study. Since students are under substantial pressure in the examination-oriented education system, it is of great importance to measure study maladaptation, which is a significant and unique source of distress among Chinese adolescents. (3) The sociality scale was only moderately correlated to the SCL-90 interpersonal dimension according to the result of concurrent validity. Our scale measures the sociality problems from the perspective of collectivist culture of China. It emphasised the relationships with parents, teachers and

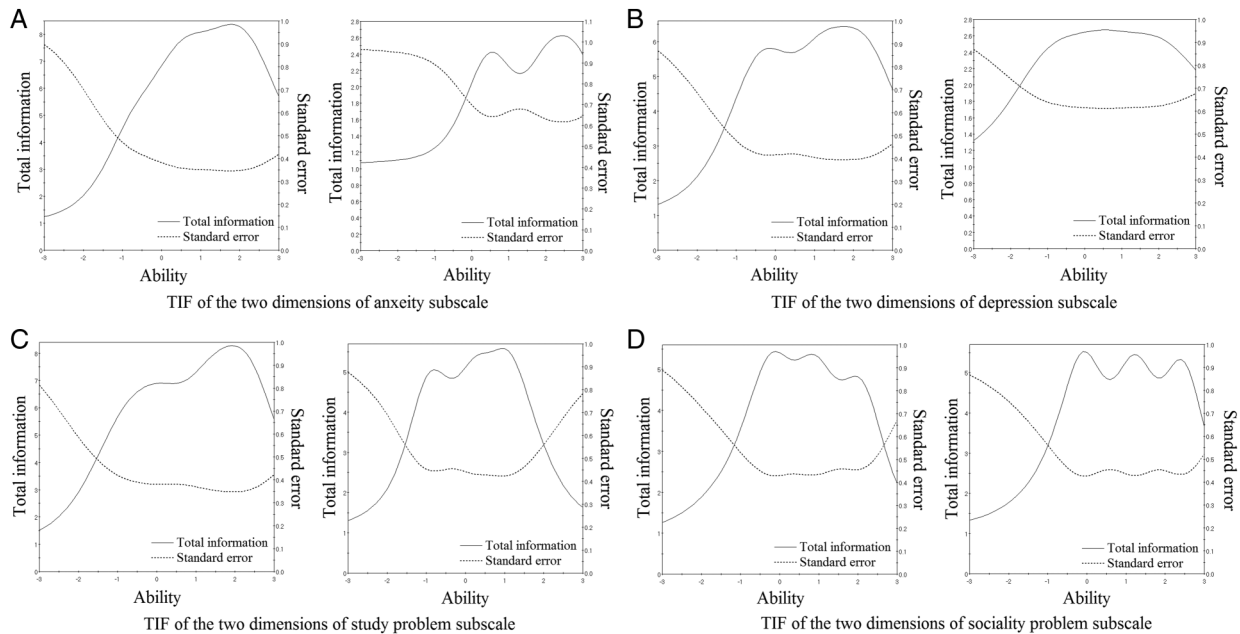


Figure 1 Test information curves. Test information curves are presented for each dimension of a scale. Ability signifies the severity of anxiety, depression, study problem and sociality problem, respectively, estimated from the item response model; it ranges from -3 to 3 . The test information of anxiety, depression and study problem peaked among students with high level of the traits, while the test information of sociality problem reached a peak among students with moderate level of the trait.

classmates. It was different from SCL-90 in that the latter measured individual feelings and thoughts that were more relevant to individualistic cultures. (4) The brief scales measure the most common mental/behavioural problems among certain populations and are time-saving.

Under the a priori hypothesis, the four scales measured different latent traits respectively. According to the EFA results, each scale had two dimensions. Dimensions of the anxiety, study and sociality scales were well explained, while the division of the depression scale dimension was psychologically obscure. CFA showed acceptable to good RMSEA and CFI for all scales, although the TLI of the sociality scale was unacceptable and local dependency was detected between a few items. Further revisions will be needed to optimise the sociality problem scale.

The scales had good internal consistency and test-retest reliability. IRT parameters showed that all items had moderate to high discriminative power, and most items had appropriate difficulty. TIF showed that all scales had strong reliability, and they were most discriminative among adolescents with moderate to high levels of emotional/behavioural problems. The peaked information function of the scales reflected the quasi-traits of psychopathology constructs. A quasi-trait refers to a unipolar construct in which one end of the scale represents severity, while the other pole represents its absence (eg, depression vs not depressed).³⁷ This is in contrast to a bipolar construct where both ends of the scale represent meaningful variation (eg, depression vs happiness, high-

health literacy vs low-health literacy), which is commonly seen in public health settings.^{45 46}

Good measurement requires that test scores have the same meaning across all relevant examinee groups. In the current study, significant gender DIF was detected at the item level, which might compromise the ability to scale boys and girls onto a common metric. In an observation on DIF research in clinical settings, McHorney and Fleishman⁴⁷ suggest that women are more likely to report physical and emotional distress as well as pain, fatigue and other marks of negative effects. However, we observed a mixed pattern of DIF between boys and girls. For the anxiety and depression scale, girls were more likely to endorse higher scores on affective symptoms, while boys were more likely to endorse somatic symptoms (rapid heartbeat). Boys were also more likely to endorse study and sociality problems. Evidences of gender difference of somatisation still remain inclusive.⁴⁸ Nevertheless, most items showed no significant DIF. Reise and Waller³⁷ suggest that the presence of item-level DIF does not necessarily lead to bias at the level of scale scores. Although the DIF of our scale was significant, the overall location parameters were basically equal between boys and girls. We considered that the observed DIF in a few items was not psychologically important, and we concluded that no large bias was observed at the level of composite.

The study has some limitations. First, owing to the large sample size and feasibility of performing structured interviews, diagnoses were not determined by psychologists/psychiatrists; as a result, the cut-off point of

the scales were not provided, and the scales could only be used to assess the level of a problem rather than to screen those with a problem. We will conduct a higher standard of validation in a further study. Second, a convenient sampling method was used to select cities and counties, considering the feasibility of field survey. Although schools and classes were randomly selected, the representativeness of the study population may be limited. Third, item-level DIF was detected, although this DIF does not necessarily lead to bias at the level of composite. To test and find DIF is better than to ignore a potential problem. In spite of the limitations, the study provides new short scales to measure common emotional and behavioural problems among adolescents in middle schools. The scales meet psychometric standards, and can serve as a reliable tool to measure the severity of common emotional, study and sociality problems among Chinese adolescents.

Contributors MH and ZS conceived and designed the study. MH collected the data. MS analysed the data and drafted the manuscript. All authors gave final approval to the version submitted for publication.

Funding This work was supported by the Natural Science Foundation of China (30400355) and China Scholarship Council (201406370034).

Competing interests None declared.

Patient consent Obtained.

Ethics approval The Ethics Committee of Central South University.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

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 and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Stormshak EA, Connell AM, Véronneau MH, *et al.* An ecological approach to promoting early adolescent mental health and social adaptation: family-centered intervention in public middle schools. *Child Dev* 2011;82:209–25.
- Davidson LL, Grigorenko EL, Boivin MJ, *et al.* A focus on adolescence to reduce neurological, mental health and substance-use disability. *Nature* 2015;527:S161–6.
- United Nations Children's Fund. Progress for Children: A Report Card on Adolescents. New York, NY: UNICEF, 2012.
- Costello EJ, Erkanli A, Angold A. Is there an epidemic of child or adolescent depression? *J Child Psychol Psychiatry* 2006;47:1263–71.
- Bennett K, Manassis K, Duda S, *et al.* Preventing child and adolescent anxiety disorders: overview of systematic reviews. *Depress Anxiety* 2015;32:909–18.
- Gore FM, Bloem PJ, Patton GC, *et al.* Global burden of disease in young people aged 10–24 years: a systematic analysis. *Lancet* 2011;377:2093–102.
- Belfer ML. Child and adolescent mental disorders: the magnitude of the problem across the globe. *J Child Psychol Psychiatry* 2008;49:226–36.
- Williams SB, O'Connor EA, Eder M, *et al.* Screening for child and adolescent depression in primary care settings: a systematic evidence review for the US Preventive Services Task Force. *Pediatrics* 2009;123:e716–35.
- Connolly SD, Bernstein GA. Work Group on Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with anxiety disorders. *J Am Acad Child Adolesc Psychiatry* 2007;46:267–83.
- Dai J, Zhong BL, Xiang YT, *et al.* Internal migration, mental health, and suicidal behaviors in young rural Chinese. *Soc Psychiatry Psychiatr Epidemiol* 2015;50:621–31.
- Qiu P, Caine E, Yang Y, *et al.* Depression and associated factors in internal migrant workers in China. *J Affect Disord* 2011;134:198–207.
- Yu Y, Liu ZW, Hu M, *et al.* Assessment of mental health literacy using a multifaceted measure among a Chinese rural population. *BMJ Open* 2015;5:e009054.
- Chen X, Chen H, Kaspar V, *et al.* Adolescent social, emotional, and school adjustment in mainland China. *Int J Group Tensions* 2000;29:51–78.
- Hofmann SG, Hinton DE. Cross-cultural aspects of anxiety disorders. *Curr Psychiatry Rep* 2014;16:450.
- Hofmann SG, Anu Asnaani MA, Hinton DE. Cultural aspects in social anxiety and social anxiety disorder. *Depress Anxiety* 2010;27:1117–27.
- Tanaka-Matsumi J. Taijin Kyofusho: diagnostic and cultural issues in Japanese psychiatry. *Cult Med Psychiatry* 1979;3:231–45.
- Chiao JY, Blizinsky KD. Culture-gene coevolution of individualism-collectivism and the serotonin transporter gene. *Proc Biol Sci* 2010;277:529–37.
- Liu J, Yan F, Ma X, *et al.* Perceptions of public attitudes towards persons with mental illness in Beijing, China: results from a representative survey. *Soc Psychiatry Psychiatr Epidemiol* 2016;51:443–53.
- Hsu L, Alden LE. Cultural influences on willingness to seek treatment for social anxiety in Chinese- and European-heritage students. *Cultur Divers Ethnic Minor Psychol* 2008;14:215–23.
- Thombs BD, Roseman M, Kloda LA. Depression screening and mental health outcomes in children and adolescents: a systematic review protocol. *Syst Rev* 2012;1:58.
- Achenbach TM, Verhulst FC, Baron GD, *et al.* A comparison of syndromes derived from the child behavior checklist for American and Dutch boys aged 6–11 and 12–16. *J Child Psychol Psychiatry* 1987;28:437–53.
- Hyler SE, Rieder RO, Williams JBW, *et al.* The personality diagnostic questionnaire: development and preliminary results. *J Pers Disord* 1998;2:229–37.
- Rutter M. A children's behaviour questionnaire for completion by teachers: preliminary findings. *J Child Psychol Psychiatry* 1967;8:1–11.
- Spence SH. A measure of anxiety symptoms among children. *Behav Res Ther* 1998;36:545–66.
- Zung WW. A rating instrument for anxiety disorders. *Psychosomatics* 1971;12:371–9.
- Zung WW. A self-rating depression scale. *Arch Gen Psychiatry* 1965;12:63–70.
- Kovacs M. The Children's Depression, Inventory (CDI). *Psychopharmacol Bull* 1985;21:995–8.
- Angold A, Costello EJ. The Child and Adolescent Psychiatric Assessment (CAPA). *J Am Acad Child Adolesc Psychiatry* 2000;39:39–48.
- Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- Liu Y, Wang J, Hinds PS, *et al.* The emotional distress of children with cancer in China: an item response analysis of C-Ped-PROMIS Anxiety and Depression short forms. *Qual Life Res* 2015;24:1491–501.
- Guo C, Tomson G, Guo J, *et al.* Psychometric evaluation of the Mental Health Continuum-Short Form (MHC-SF) in Chinese adolescents—a methodological study. *Health Qual Life Outcomes* 2015;13:198.
- Zhao J, Xing X, Wang M. Psychometric properties of the Spence Children's Anxiety Scale (SCAS) in Mainland Chinese children and adolescents. *J Anxiety Disord* 2012;26:728–36.
- Peng L, Zhang J, Li M, *et al.* Negative life events and mental health of Chinese medical students: the effect of resilience, personality and social support. *Psychiatry Res* 2012;196:138–41.
- Dai Y, Zhang B, Sun H, *et al.* Prevalence and correlates of psychological symptoms in Chinese doctors as measured with the SCL-90-R: a meta-analysis. *Res Nurs Health* 2015;38:369–83.
- Hu M, Hu G, Hu Q, *et al.* Development and validation a psychological problems screening scale for middle-school students. *Chinese J Clin Psychol* 2012;20:340–2.
- Brown TA. *Confirmatory factor analysis for applied research*. New York: The Guilford Press, 2006.
- Reise SP, Waller NG. Item response theory and clinical measurement. *Annu Rev Clin Psychol* 2009;5:27–48.

38. Thomas ML. The value of item response theory in clinical assessment: a review. *Assessment* 2011;18:291–307.
39. Baker FB. The Basics of Item Response Theory (second edition). *United States of America: ERIC Clearinghouse on Assessment and Evaluation*; 2001.
40. Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: applications of an EM algorithm. *Psychometrika* 1981;46:443–59.
41. von Elm E, Altman DG, Egger M, *et al.* STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370:1453–7.
42. Hsu LK, Folstein MF. Somatoform disorders in Caucasian and Chinese Americans. *J Nerv Ment Dis* 1997;185:382–7.
43. Kirmayer LJ, Sartorius N. Cultural models and somatic syndromes. *Psychosom Med* 2007;69:832–40.
44. Waza K, Graham AV, Zyzanski SJ, *et al.* Comparison of symptoms in Japanese and American depressed primary care patients. *Fam Pract* 1999;16:528–33.
45. Shen M, Hu M, Liu S, *et al.* Assessment of the Chinese Resident Health Literacy Scale in a population-based sample in South China. *BMC Public Health* 2015;15:637.
46. Shen M, Hu M, Sun Z. Assessment of school-based quasi-experimental nutrition and food safety health education for primary school students in two poverty-stricken counties of west China. *PLoS ONE* 2015;10:e0145090.
47. McHorney CA, Fleishman JA. Assessing and understanding measurement equivalence in health outcome measures: issues for further quantitative and qualitative inquiry. *Med Care* 2006;44(Suppl 3):S205–10.
48. Piccinelli M, Simon G. Gender and cross-cultural differences in somatic symptoms associated with emotional distress. An international study in primary care. *Psychol Med* 1997;27:433–44.