Zhou et al. BMC Psychiatry (2020) 20:480
https://doi.org/10.1186/s12888-020-02859-8

BMC Psychiatry

RESEARCH ARTICLE Open Access

Are comparisons of mental disorders between Chinese and German students possible? An examination of measurement invariance for the PHQ-15, PHQ-9 and GAD-7

Yan Zhou[1*], Jing Xu[2] and Winfried Rief[1]

Abstract

Background: The Patient Health Questionnaire (PHQ) is one of the most commonly used instruments to assess mental disorders. However, research on its cross-cultural measurement invariance is not yet sufficient. This study examined the measurement invariance of the Chinese and German versions of the PHQ's somatic symptom severity scale (PHQ-15), depressive symptom severity scale (PHQ-9) and seven-item Generalized Anxiety Disorder (GAD-7) scale as a prerequisite for their use in cross-cultural comparisons.

Methods: We used online data collected from groups of Chinese students in China ($n = 413$) and German students in Germany ($n = 416$). Separate measurement models for each group were examined using confirmatory factor analysis and measurement invariance testing was conducted to test the cross-cultural equivalence.

Results: Findings demonstrated that the PHQ-9 and GAD-7 had partial scalar measurement invariance, but the cross-cultural measurement invariance of the PHQ-15 could not be confirmed. Comparisons of latent means did not indicate differences in the levels of depression and anxiety symptoms between Chinese and German samples.

Conclusion: The PHQ-9 and GAD-7 can be used in cross-cultural comparison of prevalence, but the intercultural use of PHQ-15 is more problematic. Findings are discussed from intercultural and methodological perspectives.

Keywords: Cross-cultural comparison, Measurement invariance, Patient Health Questionnaire-15 (PHQ-15), Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7 (GAD-7)

Background

Depression, anxiety disorders and somatoform disorders are the most common mental disorders worldwide and differences in epidemiology exist across countries and cultures [1, 2]. Previous studies showed that base rates of depression and anxiety disorders are lower in China than in American and European countries [1, 3–5], and prevalence rates of somatoform disorders are inconsistent [6, 7]. For example, the 12-month prevalence of anxiety disorders in China was 5.0% compared to 15.3% in Germany [8, 9] and for major depressive episodes was 3.6% in China and 6% in Germany [8, 10]. Cultural, linguistic and methodological aspects could contribute to explaining the differences in prevalence rates of disorders. According to a literature review [4], the lower prevalence of major depressive disorders that persisted in East/Southeast Asia compared to other regions of the world still remained, even after adjusting for methodological differences. The study showed evidence that cross-national differences may reflect either true prevalence differences or the cross-cultural insensitivity of

* Correspondence: zhouyan.1.zhouyan@gmail.com
[1]Clinical Psychology and Psychotherapy, Department of Psychology, Philipps-University of Marburg, Gutenbergstr. 18, D-35032 Marburg, Germany
Full list of author information is available at the end of the article

diagnostic criteria such as the *Diagnostic and Statistical Manual of Mental Disorders* [11] (DSM) and the *International Classification of Diseases* [12] (ICD) [13]. A deeper cross-cultural examination of these differences is overdue.

The Patient Health Questionnaire [14], which includes the somatic symptom severity scale (PHQ-15), the depressive symptom severity scale (PHQ-9) and the seven-item Generalized Anxiety Disorder (GAD-7) scale, screens, identifies and measures the severity of most common mental disorders and is one of the most commonly used instruments to assess psychological disorders. It is a short self-report questionnaire based on the diagnostic criteria of the DSM-IV and its scales also have a high level of suitability for the newly developed DSM-V [11], with the American Psychiatric Association (APA) recommending them for measuring the severity of depression, anxiety disorders and somatoform disorders [15]. In both Western and Chinese general populations it showed good reliability and validity of scores [16–22]. Furthermore, taking into consideration that Western psychologization is stronger than Chinese somatization, this self-completed questionnaire had the advantage of revealing more psychological distress than in interviews with the Chinese population [23].

In order to accurately compare the cross-cultural prevalence of these mental disorders, diagnostic measurements such as the PHQ must be measurement invariant across cultures and languages (comparable both cross-culturally and cross-linguistically). DSM- or ICD-based diagnostic measures were criticized as not being culturally sensitive enough due to culture-specific forms of disorders [24]. Cultural differences in scale scores can also result from differences in the understanding of certain concepts, translation problems, frequency of word use or other measurement errors, and the potentially biased items should be identified before comparison [25]. Despite the widespread use of the PHQ, cross-cultural examination of its measurement invariance has been mostly neglected and such examinations have scarcely been made between mainland Chinese and Western samples.

The commonly used measurement equivalence approach (also called measurement invariance) is confirmatory factory analysis (CFA) and this offers a robust statistical framework for testing measurement invariance. The most frequently assessed levels of measurement invariance included configural, metric and scalar invariance, which refer to different model parameters of a measurement model and build on each other in a hierarchical structure. Configural invariance, the least restrictive form of measurement invariance, is present when the number of factors and the pattern of the factor loadings between the latent variables and indicators in the compared groups are similar. When the factor loadings of items are also invariant across groups, then metric invariance could be assumed. Scalar invariance exists if, additionally, the intercepts of the indicator variables are identical [26]. Scalar invariance or at least partial measurement invariance, which is declared when at least two items per latent variable (i.e., factor loadings, factor intercepts) are found to be invariant, is a prerequisite for the comparison of latent mean values obtained from subsamples [27–29].

In previous studies, measurement invariance of the PHQ-15 with a bifactor model (one general somatic symptom factor and four orthogonal-specific symptom factors of pain, gastroenterology, cardiovascular and fatigue symptoms) could be confirmed with samples of college students from Germany and Switzerland [30], between German and migrants [31] and between patient samples from Germany and the Netherlands, but not between Chinese patient samples and Western (German and Dutch) patient samples [32]. Measurement invariance with a one-factor model was confirmed between primary care patients of native-born Germans, Russian-speaking immigrants and native-born Russians [33]. So far, hardly any studies have explored measurement invariance of the PHQ-15 in samples from mainland China and Western countries.

Previous studies have predominantly been able to confirm a one-factor structure of the PHQ-9 with different samples across cultures or migrants in Western countries and found measurement invariance of the scale in six ethnic groups in the Netherlands, in diverse college populations in the USA and in first- and second-generation migrants of the German population [34–38]. However, the items “sleep problems”, “appetite changes” and “anhedonia” showed cross-cultural measurement biases when comparing Turkish immigrants and Germans, and the item “psychomotor problems” seemed to be culturally biased in Surinam Dutch males compared to Dutch males. A bifactor structure (“somatic factor” and “non-somatic factor”) could be confirmed in a German study with samples across gender [39] and in a Japanese study between clinical and nonclinical samples [40].

Existing evidence demonstrates that the GAD-7 has good psychometric properties and shows reliability and validity of scores as a measure of anxiety in the German general population [19] and in various samples in Chinese primary care [21, 41]. So far, only a few studies have examined cross-cultural measurement invariance of the GAD-7. A study by Parkerson, Thibodeau, Brandt, Zvolensky, and Asmundson [42] has confirmed a revised unitary model of the GAD-7 and found that Black/African Americans with high GAD symptoms scored lower on the GAD-7 than White/Caucasian Hispanic participants. It indicated that the GAD-7 is not culturally sensitive enough and the lower prevalence rate for the

Black/African American sample could reflect cross-cultural measurement biases related to the diagnostic instrument rather than true differences in GAD symptoms. It is still unclear whether such cross-cultural measurement biases also exist in the epidemiological comparison of cultural groups from China and Western countries.

To address the above-mentioned limitations of current studies in examining cross-cultural measurement biases of the PHQ in samples from mainland China and Western countries, we conducted this study to investigate measurement invariance of the PHQ-15, PHQ-9 and GAD-7 across Chinese and Western (represented by Germany) cultures. We investigated student samples because of the advantage of comparability in educational status, age and other psychosocial aspects, but also the different cultural backgrounds. Based on previous research, we expected there to be measurement invariance of the PHQ-9 between the two cultures but predicting the same for the PHQ-15 may be problematic. Due to lack of previous studies, we did not make a hypothesis about the intercultural measurement invariance of the GAD-7. Differences in latent means for somatic symptoms, depression and anxiety syndromes were also assessed if scalar measurement invariance across cultures was demonstrated. Investigating the cross-cultural equivalence of the PHQ-15, PHQ-9 and GAD-7 has high relevance to the diagnosis of mental disorders and is a prerequisite for cross-cultural comparisons.

Methods

Participants

The online data used in the present study are from a dataset collected in a project for intercultural comparison of willingness to seek psychological help [43]. The data were collected in Germany and China in August 2016 and the collection lasted for 6 months. German students at the University of Marburg (total number of students: 26,355) were invited to take part in the survey via the university email list. To increase the interest to participate, they could be entered into a draw for vouchers worth 20 euros. Chinese students at the University of Zhengzhou in China were recruited on "WeChat", a popular social media platform used by most Chinese students, and they received no financial reward. Chinese students who were in the WeChat groups of various affiliated faculties (e.g., Economics and Electronic Information Engineering: 9156 students) were invited to participate in this study. After the application of exclusion criteria (no migration background; a minimum scale processing time of 10 min), the Chinese sample available for analysis decreased from 566 to 413 and the German sample from 456 to 416. The demographic characteristics of the two groups are summarized in Table 1.[1] The study was approved by the ethics committee of the Faculty of Psychology at the University of Marburg (approval number: 2016-19 k).

Assessment instruments

PHQ-15

The PHQ-15 was used to assess and diagnose somatoform disorders [44] and includes 15 prevalent somatic symptoms that represent the most common symptoms observed in primary care that typically cannot be fully explained by a diagnosed general medical condition. Two of the items were from the depression subscale of the PHQ-9 ("*Trouble falling or staying asleep, or sleeping too much*"; "*Feeling tired or having little energy*"). Three response categories were offered: "not bothered at all", "bothered a little" or "bothered a lot". The total score ranged from 0 to 30. The reliability and validity of the scores were supported by studies both in German and Chinese populations [17, 18, 22].

PHQ-9

The PHQ-9 was used to assess and diagnose depression [45]. The participants responded on a four-point Likert scale and the total score ranged from 0 to 27. The PHQ-9 has good psychometric properties and includes high sensitivity for depressive disorders and good specificity for screening of patients with depression in both Chinese and German general populations [20, 46] and in their corresponding primary care populations [16, 17]. The PHQ-9 was considered superior to other self-rating instruments for the detection of depressive disorders [17].

GAD-7

The seven-item GAD-7 was developed to identify potential patients with a generalized anxiety disorder [47] and to assess the severity of symptoms of general anxiety because of its good operating characteristics for anxiety disorders [48]. Participants indicated agreement with the presence of symptoms such as "*Feeling nervous, anxious or on edge*" and "*Not being able to stop or control worrying*" on a four-point Likert scale ranging from 0 (*not at all*) to 3 (*nearly every day*). The total score ranges from 0 to 21.

[1] "Since the demographic data of the two samples was significantly different, propensity score matching (by gender and age) was performed to exclude the test effect of gender and age for the purpose of data analysis. However, the results did not differ from those obtained prior to performing the propensity score matching test. The results of the comparison between the underlying structure and the invariant items of the scales remained unchanged. Therefore, we presented the results without conducting propensity score matching, owing to the large sample size" [43].

Table 1 Participant Demographic Characteristics

Variables	German Students ($n = 416$)	Chinese Students ($n = 413$)	*p*-value	Effect size
Sex [female] (%)	294 (70.7%)	250 (60.5%)	.001	$\varphi = .15$
Age (Mean ± SD; range)	23.85 ± 4.66; 18–58	20.79 ± 2.57; 17–39	$p < .001$	Cohen's $d = .81$
Current Academic Degree			p < .001	Cramer's V = .78
Bachelor (%)	183 (44%)	391 (94.7%)		
Master (%)	83 (20%)	7 (1.7%)		
Ph.D. (%)	17 (4.1%)	4 (1.0%)		
Others (%)	137 (32%)	11 (2.7%)		

Note. "Other" refers to a combined Bachelor's and Master's degree program. The comparisons of sex and academic degree were calculated using Chi-squared test. The age comparison was calculated using T-test

Translation

The German validated versions of the PHQ-15, PHQ-9 and GAD-7 [14] were used in the German sample and the translation was done according to "state of the art criteria" using the translation/retranslation method. The Chinese versions of the PHQ-15 [22], PHQ-9 [46] and GAD-7 [49] were also validated in previous studies and the translation followed the customary translation/back-translation method.

Statistical analysis

First, SPSS (version 25, IBM, Armonk, USA) software was used for checking the descriptive statistics (means, standard deviations, skewness and kurtosis of the sum scores and evidence of internal consistencies for each scale and each sample), and then we used the software program M*plus* v7.4 [50] for further data analysis. We examined separate measurement models for each group using confirmatory factor analysis (CFA). To assess the model fit we used χ^2 difference tests, as recommended by Hu and Bentler [51]. Because the χ^2 difference test is sensitive to sample size, other common indices to assess the goodness of model fit were also used: comparative fit index (CFI), root mean square error of approximation (RMSEA), standardized root mean residual (SRMS) and difference in CFI between the base model and the constrained model (ΔCFI). The following cutoff values were used: CFI ≥ .90 [52], RMSEA ≤ .08 and SRMR ≤ .08 [53].

Then the step-up approach was applied to add a series of increasingly stringent equality constraints to the models [27, 54]. The configural invariance of the baseline model was estimated as the starting point of the multiple group comparisons, in which all parameters (factor loadings and intercepts of indicators) vary freely. We investigated whether the construct was similarly displayed in different groups, meaning that both the number of specified factors and the indicators that load on the factors should be comparable. In the next step, the metric invariance was checked by constraining the factor loadings of indicators to be equal. Then the scalar invariance, the next highest form of measurement invariance, was assessed by additionally constraining intercepts of indicators to be equal. After gradual equality constraining of the parameters across the groups, these models were compared with the baseline model. The decision on whether a model was accepted or not was made according to the χ^2 difference test [51]. As the χ^2 value was dependent on sample size, Cheung and Rensvold [55] suggested that the difference in CFI between the baseline model and the constrained model should not be more than 0.01. If the full measurement invariance cannot be confirmed, partial invariance should be examined [28]. The constrained model based on the modification indices was subsequently examined by releasing the equality constraints in descending order for misspecified items. At least two loadings or intercepts should be equal between groups in order to establish partial measurement invariance. If evidence for scalar invariance or at least partial scalar invariance[2] exists, then the latent means of samples could be compared [27, 28].

Results

Descriptive statistics

Means, standard deviations, skewness and kurtosis of the sum scores and evidence of internal consistencies for each scale and each sample are presented in Table 2. According to the cutoff values (skewness ≤3, kurtosis ≤8) recommended by Kline (2010), skewness and kurtosis indicated a normal distribution in the samples. The internal consistency of the scores was at least good (> .70). Items 2 ("back pain") and 9 ("fainting spells") in the German version of the PHQ-15 showed a small correlation (< .10) with other items of the scale, mainly because of

[2]Partial measurement invariance is declared when at least two items per latent variable (i.e., factor loadings, factor intercepts) are found to be invariant [29].

Table 2 Means, Standard Deviations, Skewness, Kurtosis, and Internal Consistency across Scales and Samples

	German Students					Chinese Students				
Scale	*M*	*SD*	Skew	Kurt	α	*M*	*SD*	Skew	Kurt	α
PHQ-15	7.39	4.49	.77	.44	.76	6.87	4.67	.75	.22	.83
PHQ-9	6.77	4.84	.76	.09	.85	6.99	4.76	1.01	1.49	.88
GAD-7	6.23	4.27	.80	.06	.87	5.38	4.00	1.18	1.93	.90

Notes. PHQ-15 = Patient Health Questionnaire-15 Physical Symptoms; PHQ-9 = Patient Health Questionnaire-9 Depression Symptoms;
GAD-7 = Generalized Anxiety
Disorder 7-item Scale; Skew = Skewness; Kurt = Kurtosis

very low or very high base rates compared to other symptoms. Despite these findings, we first tested the original models using CFA.

Measurement invariance of the PHQ-15

Single-group CFA

Results from CFA are presented in Table 3. The unidimensional model of the PHQ-15 was examined first, which assumes only one latent factor (model A). In both groups, the PHQ-15 resulted in acceptable SRMR but poor CFI and RMSEA values (Chinese group: CFI = .827, RMSEA = .079, 90% CI [.070, .089], SRMR = .057; German group: CFI = .716, RMSEA = .086, 90% CI [.077, .096], SRMR = .065). This means that a one-factor solution does not fit the samples of Chinese and German students. Then we tried the hierarchical measurement model with four first-order latent factors and a second-order latent factor (model B) recommended by Mewes et al. [31], which was based on the criteria for somatoform disorders and physical complaints of depressive disorders in ICD-10 and DSM-IV. The four factors are as follows: pain symptoms (item 2 "back pain," item 3 "pain in your arms, legs or joints," item 4 "menstrual cramps or other problems with your periods", item 5 "pain or problems during sexual intercourse", item 6 "headaches"), gastrointestinal symptoms (item 1 "stomach pain", item 12 "constipation, loose bowels or diarrhea", item 13 "nausea, gas or indigestion"), cardiovascular symptoms (item 7 "chest pain", item 8 "dizziness", item 9 "fainting spells", item 10 "feeling your heart pound or race", item 11 "shortness of breath") and fatigue symptoms (item 14 "trouble sleeping", item 15 "feeling tired or having low energy") (see Supplementary Material, Table S1). The model with four first-order latent factors and a second-order latent factor achieved an acceptable fit for both samples in terms of RMSEA (Chinese group: CFI = .936, RMSEA = .050, 90% CI [.039, .061], SRMR = .042; German group: CFI = .914,

Table 3 Fit Indices from Comparative Factor Analysis (CFA) and Invariance Analyses between Groups for the PHQ-15

	χ2 *(df)*	CFI	RMSEA [90% CI]	SRMR	ΔCFI	Δχ2 *(df)*
model A. one-factor model						
German Students	369.242 (90)	.716	.086 [.077, .096]	.065		
Chinese Students	323.274 (90)	.827	.079 [.070, .089]	.057		
model B. four-factor model						
German Students	168.941 (84)	.914	.049 [.038, .060]	.049	.016	200.301 (6)
Chinese Students	170.681 (84)	.936	.050 [.039, .061]	.042	.015	152.593 (6)
Multiple group CFA model B						
Configural invariance	339.622 (168)	.926	.050 [.042, .057]	.045		
Metric invariance	410.344 (183)	.902	.055 [.048, .062]	.071	.024	70.722 (15)
λ_9 free	393.042 (182)	.909	.053 [.046, .060]	.063	.023	53.420 (14)
λ_9, λ_{10} free	383.574 (181)	.913	.052 [.045, .059]	.059	.013	43.952 (13)
λ_9, λ_{10}, λ_{11} free	371.885 (180)	.918	.051 [.043, .058]	.056	.008	32.263 (12), $p > .001$
Scalar invariance	556.111 (195)	.845	.067 [.060, .073]	.067	.073	184.226 (15)
τ_{10} free	501.412 (194)	.868	.062 [.059, .069]	.062	.050	129.527 (14)
τ_{10}, τ_2 free	482.435 (193)	.876	.060 [.053, .067]	.060	.042	110.55 (13)
τ_{10}, τ_2, τ_5 free	469.157 (192)	.881	.059 [.052, .066]	.059	.037	97.272 (12)
τ_{10}, τ_2, τ_5, τ_{12} free	451.368 (191)	.887	.058 [.051, .065]	.058	.031	79.483 (11)
τ_{10}, τ_2, τ_5, τ_{12}, τ_9 free	440.687 (190)	.892	.056 [.050, .063]	.056	.026	68.802 (10)
τ_{10}, τ_2, τ_5, τ_{12}, τ_9, τ_6 free	430.746 (189)	.896	.056 [.049, .062]	.054	.022	58.864 (9), $p < .001$

Notes. PHQ-15 = Patient Health Questionnaire-15 Physical Symptoms; Item 2 = 'back pain'; item 5 = 'pain or problems during sexual intercourse'; item 6 = 'headaches'; item 9 = 'fainting spells'; item 10 = 'feeling your heart pound or race'; item 11 = 'shortness of breath'; item 12 = 'constipation, loose bowels, or diarrhea.' All χ^2 tests and Δχ2 were significant, $p < .001$

RMSEA = .049, 90% CI [.038, .060], SRMR = .049). Baseline models for analysis of measurement invariance between cultures could be established.

Measurement invariance between cultures

After confirming the superiority of model B compared to model A, measurement invariance analysis between cultures was performed. The testing results of measurement invariance for the PHQ-15 are presented in Table 3. The baseline model of the PHQ-15 showed acceptable configural invariance (CFI = .926, RMSEA = .050, 90% CI [.042, .057], SRMR = .045). In the next step, the metric invariance was tested by constraining the factor loadings to be equal. The fit of the metric invariance was poor, with a decrease in CFI of more than 0.01 (CFI = .902, RMSEA = .055, 90% CI [.048, .062], SRMR = .071, ΔCFI = .024). Modification indices indicated that the loading of items 9 (fainting spells), 10 (feeling your heart pound or race) and 11 (shortness of breath) differed across the groups. After releasing the constraints for these items in descending order, the fit of this modified model was acceptable (CFI = .918, RMSEA = .051, 90% CI [.043, .058], SRMR = .056, ΔCFI = .008). Then the factor intercepts were constrained to be equal and the scalar invariance was shown to be poor, with a CFI of .845 and a drop in CFI of more than 0.01 (ΔCFI = .073). The modification indices showed that the intercepts of items between the two groups were invariant. After releasing the equality constraints for items 10 (feeling your heart pound or race), 2 (back pain), 5 (pain or problems during sexual intercourse), 12 (trouble sleeping), 9 (fainting spells) and 6 (headaches) in descending order, the fit of this modified model for checking partial scalar invariance was still unacceptable, with a poor model fit and a drop in CFI of more than 0.01 (CFI = .897, ΔCFI = .021). Hence, the partial scalar invariance of the four-factor model between the groups could not be established and comparison of the latent means could not be conducted.

Measurement invariance of the PHQ-9

Single-group CFA

Similar to the PHQ-15, we first examined the fit of the one-factor model of the PHQ-9 using CFA in the two groups (Table 4). The one-factor model of the PHQ-9 had acceptable CFI and SRMR in both groups but poor RMSEA, with values of more than .08 (Chinese group: CFI = .951, RMSEA = .082, 90% CI [.065, .099], SRMR = .038; German group: CFI = .900, RMSEA = .104, 90% CI [.088, .120], SRMR = .051). Therefore we tried a two-factor-solution, which was suggested by Petersen et al. [39]. The two factors included "somatic" (e.g., sleep disturbances, fatigue and appetite changes) and "non-somatic" items (e.g., depressed mood, lack of interest and suicidal ideation). The model with two latent factors afforded a good fit in both samples (Chinese group: CFI = .957, RMSEA = .078, 90% CI [.061, .096], SRMR = .037; German group: CFI = .969, RMSEA = .059, 90% CI [.040, .077], SRMR = .033). A baseline model for analysis of measurement invariance between the two groups could be established.

Table 4 Fit Indices from Comparative Factor Analysis (CFA) and Invariance Analyses between Groups for the PHQ-9

	χ2 *(df)*	CFI	RMSEA [90% CI]	SRMR	ΔCFI	Δχ2 *(df)*
model A. one-factor model						
German Students	147.851 (27)	.900	.104 [.088, .120]	.051		
Chinese Students	101.362 (27)	.951	.082 [.065, .099]	.038		
model B. two-factor model						
German Students	63.070 (26)	969	.059 [.040, .077]	.033		84.781
Chinese Students	91.453 (26)	.957	.078 [.061, .096]	.037		9.909, p > .001
Multiple groups CFA models						
Configural invariance	154.523 (52)	.962	.069 [.057, .082]	.035		
Metric invariance	229.833 (61)	.938	.082 [.071, .093]	.082	.024	75.310 (9)
λ_8 free	185.565 (60)	.954	.071 [.060, .083]	.064	.008	31.042 (8)
λ_8, λ_1 free	174.288 (59)	.958	.069 [.057, .081]	.068	.004	19.765 (7)
λ_8, λ_1, λ_3 free	169.246 (58)	.959	.068 [.056, .080]	.061	.002	14.723 (6), p > .001
Scale invariance	302.171 (65)	.913	.094 [.083, .105]	.085	.046	132.925 (7)
τ_8 free	216.478 (64)	.944	.076 [.065, .087]	.063	.015	47.232(6)
τ_8, τ_4 free	191.586 (63)	.953	.070 [.059, .082]	.061	.006	22.340 (5)
τ_8, τ_4, τ_1 free	179.934 (62)	.957	.068 [.056, .079]	.061	.002	10.688 (4), p > .001

Note. PHQ-9 = Patient Health Questionnaire-9 Depression Symptoms; Item 1 = 'lack of interest'; item 3 = 'sleep difficulties'; item 4 = 'feeling tired or having little energy'; item 8 = 'moving or speaking slowly, or fretful.' All χ2 tests and Δχ2 were significant, p < .001

Measurement invariance between cultures

We examine the measurement invariance across cultures with model B because of its better model fit than model A. The model specifications for the PHQ-9 are displayed in Table 4. The global fit for the configural model was good (CFI = .962, RMSEA = .069, 90% CI [.057, .082], SRMR = .035). Then, item loadings were constrained to be equal in the metric invariance model. The global fit was poor, with RMSEA and SRMR bigger than .080 and ΔCFI bigger than .01. Modification indicated that loadings of items 8, 1 and 3 were invariant. The loading of items 1 (lack of interest) and 8 (moving or speaking slowly, or fretful) was higher in the Chinese sample and for item 3 (sleep difficulties) was higher in the German sample. Partial metric invariance was established by allowing the loadings of these items to vary in descending order (CFI = .959, RMSEA = .068, 90% CI [.056, .080], SRMR = .061, ΔCFI = .002). At the level of scalar invariance, RMSEA and SRMR were also bigger than .08 and the drop in CFI was larger than .01. After releasing the equality constraints of items 8, 4, and 1 in descending order, partial scalar invariance could be established (CFI = .957, RMSEA = .068, 90% CI [.056, .079], SRMR = .061, ΔCFI = .002).

Latent mean comparison

Comparison of the latent means was based on five invariant items (items 2, 5, 6, 7 and 9) and the German sample was used as the reference group. The Chinese students had a higher latent mean than German students, which means that Chinese students have more depressive symptoms than German students, but the mean difference was not significant ($z = .344$, $d = .153$, $p = .365$).

Measurement invariance of the GAD-7

Single-group CFA

CFA of the original one-factor model demonstrated an acceptable global fit in the sample of Chinese students, but the RMSEA indicated a poor fit in the sample of German students (Table 5). Modification indices suggested that the error terms of items 5 ("being so restless that it is hard to sit still") and 6 ("becoming easily annoyed or irritated") were correlated in both samples, which was similar to the findings from Parkerson et al. (2015). To improve the comparability of the two groups, correlation between the two item errors was allowed and this produced an acceptable RMSEA for the sample of German students. At the same time, the global model fit for the sample of Chinese students was also improved significantly ($\Delta\chi^2$ (df) = 15.219 (1), $p < .001$).

Measurement invariance between cultures

The results of tests of the measurement invariance of the GAD-7 are presented in Table 5. The baseline model of the GAD-7 demonstrated a good global fit (CFI = .978, RMSEA = .074, 90% CI [.057, .091], SRMR = .030) and its configural invariance was confirmed. At the level of metric invariance, the RMSEA was larger than .08 and the drop in CFI was larger than .01 (RMSEA = .081, 90% CI [.066, .097], ΔCFI = .012). Modification indices indicated that the loading of item 1 was not invariant. The loading of item 1 was higher in the German sample than in the Chinese sample. A modified model by releasing the equality constraints for item 1 provided a good fit and the assumption of metric invariance held (CFI = .973, RMSEA = .074, 90% CI [.059, .090], SRMR = .048, ΔCFI = .005). On testing for scalar invariance, the RMSEA was larger than .08 and the drop in CFI was larger than .01 (RMSEA = .095, 90% CI [.082, .109], ΔCFI = .026). Modification indices showed that the intercepts of items 4, 1 and 2 were higher in the German sample than the Chinese sample. By releasing the equality constraints of items 4, 1 and 2 in descending order, the global fit of this model was improved (CFI = .969, RMSEA = .075, 90% CI [.061, .090], SRMR = .052, ΔCFI = .004) and partial scalar invariance was established.

Latent mean comparison

Comparison of the latent means was based on four invariant items (items 3, 5, 6 and 7) and the sample of German students was used as the reference group. The Chinese students had a lower latent mean than German students on the GAD-7, but the difference was not significant ($z = -.023$, $d = .023$, $p = .759$).

Discussion

In our study, we examined the cross-cultural measurement invariance of the PHQ-15, PHQ-9 and GAD-7 by comparing two cultural groups of students, one from mainland China and the other from Germany. The results demonstrated that the original one-factor model of the PHQ-15 fitted neither of the groups. The bifactor model (one general factor and four orthogonal symptom-specific factors) of the PHQ-15 showed a better model fit in both groups but only configural and metric invariance between the groups could be confirmed, therefore it is not recommended for the cross-cultural comparison of means. The PHQ-9 and GAD-7 had the same factor structure in the two groups and showed partial scalar invariance. This means that although these scales show differences on individual items, they are generally comparable across the two cultural groups of students, which provides the possibility for cross-cultural comparative studies in the future.

We could not confirm the bifactor model (one general factor and four orthogonal symptom-specific factors) of the PHQ-15 with the cross-cultural student samples as suggested by Mewes et al. [31]. We also did not find full

Table 5 Fit Indices from Comparative Factor Analysis (CFA) and Invariance Analyses between Groups for the GAD-7

	χ2 *(df)*	CFI	RMSEA [90% CI]	SRMR	ΔCFI	Δχ2 *(df)*
Single group CFA – original one-factor model						
German Students	53.671 (14)	.967	.083 [.060, .106]	.035		
Chinese Students	49.414 (14)	.977	.078 [.055, .102]	.028		
Single group CFA ($\theta_{5,\ 6}$ free)						
German Students	44.662 (13)	.974	.077 [.053, .102]	.031	.013	9.009, $p > .001$
Chinese Students	34.195 (13)	.986	.063 [.038, .089]	.023	.009	15.219
Multiple group CFA models ($\theta_{5,\ 6}$ free)						
Configural invariance	87.866 (27)	.978	.074 [.057, .091]	.030		
Metric invariance	127.041 (34)	.966	.081 [.066, .097]	.073	.012	39.175 (7)
λ_1 free	107.649 (33)	.973	.074 [.059, .090]	.048	.005	19.783 (6), p > .001
Scale invariance	185.052 (39)	.947	.095 [.082, .109]	.063	.026	77.403 (6)
τ_4 free	164.669 (38)	.954	.090 [.076, .104]	.058	.019	57.020 (5)
τ_4, τ_1 free	137.262 (37)	.963	.081 [.067, .096]	.052	.010	29.613 (4)
τ_4, τ_1, τ_2 free	120.930 (36)	.969	.075 [.061, .090]	.052	.004	13.281 (3), p > .001

Note. GAD-7 = Generalized Anxiety Disorder 7-item; Item 1 = 'feeling nervous, anxious, or on edge'; item 2 = 'not being able to stop or control worrying'; item 4 = 'trouble relaxing.' All χ2 tests and Δχ2 were significant, $p < .001$

metric and partial scalar invariance. The possible reason for this could be that the samples included in our study have a greater difference in cultural background. Our result corresponded with the findings of an earlier cross-cultural study [32], which also could not confirm measurement invariance of the PHQ-15 between Chinese and German samples of outpatients. In our study, the pattern of variant items at the level of metric and scalar invariance across groups was mixed. Chinese students are more likely to endorse items 10 ("feeling your heart pound or race"), 11 ("shortness of breath"), 9 ("fainting spells") and 12 ("constipation, loose bowels, or diarrhea") and German students are more likely to endorse items 5 ("pain or problems during sexual intercourse"), 6 ("headaches") and 2 ("back pain"). Regarding the differences between individual items, there was a slight attempt in previous studies to focus on the influence of culture on shaping somatic awareness. A possible explanation for the differences could be that the levels of interoceptive accuracy and somatic awareness between people from Western and non-Western countries are different [56], and this phenomenon could be more strongly expressed on certain somatic symptoms in cross-cultural comparisons. Somatic awareness is a top-down process that is driven by attention, beliefs and expectations [57, 58] and these factors may affect the evaluation of the importance of different physical symptoms in different cultures. Linguistics is an important approach for studying this cultural difference. For example, future research could focus on whether certain body parts are used more than others in the description of physical states in the Chinese and German languages. In terms of methodology for testing a series of equality constraints on parameters in measurement models such as the PHQ-15 that have a complex structure across groups, multi-group CFA has the limitation that "the standard model fit criteria do not represent 'golden rules'" [59]. An alternative approach could be the multi-group exploratory structural equation modeling recommended by Marsh et al. [60], which can test measurement invariance directly and is viable for scales with a complex structure.

Consistent with the results of previous studies by Doi et al. [61] and Petersen et al. [39], a bifactor structure of the PHQ-9 could be confirmed in our study. We found partial metric and partial scalar invariance of the PHQ-9 across the two cultural groups. Chinese students are more likely to endorse items 1 (lack of interest) and 8 (moving or speaking slowly, or being fidgety) and German students are more likely to endorse items 3 (sleep difficulties) and 4 (fatigue). The higher score on item 1 (lack of interest) is consistent with the results of the study by Leung [62], which found that East Asian students who share the Confucian culture (high regard for academic achievement) displayed relatively negative attitudes toward learning even though they outperformed Western students. Hau and Ho [63] have reviewed the previous studies and found that Chinese students are more likely to study under external pressure and have lower interest in studying. Regarding "sleep problems", our study could support Parker, Cheah and Roy's [64] finding that insomnia is not being overrepresented in the Chinese sample, although many Asian psychiatrists have seen it as one of the most common reasons for depressed Chinese to seek help. It appears to be a true concomitant of depression and not distinctly culturally determined.

Chinese students may have a higher prevalence of depression than other populations in China because they are more open and inclined to express emotional distress. This is in line with the comparison of latent means of the two groups, showing that German students did not express more depressive symptoms than Chinese students, although previous studies have found that the prevalence of depression disorders was lower in Southeast Asia (including China) than in Western Europe [4, 65]. To use the PHQ-9 in the general Chinese population, who are not necessarily willing to report the emotional symptoms of depression or are less aware of them, a lower cut-off value would be advisable in order to maximize the detection of people with depression [66].

Partial scalar invariance of the original one-factor model of the GAD-7 could be confirmed across groups with Chinese and German students. The difference across groups indicated that German students are more likely to report anxiety symptoms such as "feeling anxious" (item 1), "not being able to stop worrying" (item 2) and "trouble relaxing" (item 4). But these differences were not significant and the latent means of the two groups did not differ, which means that German students did not have significantly higher levels of general anxiety than Chinese students. This is not consistent with the results of previous studies, which show that non-Western cultures have less risk of anxiety disorder [3, 67]. In Asian countries, culture-specific anxiety symptoms such as shame [68] were not included in the GAD-7 and it is unclear whether such aspects play a role in the measurement of general anxiety severity because empirical research is lacking.

Limitations

This study has some limitations that should be considered. First, our study was conducted in samples of college students, which controlled for other non-cultural factors contributing to the results (e.g., education), but it is unclear whether the findings of this study can be generalized to other population groups. It could be more difficult to establish measurement invariance in other populations across cultures because the younger generation who grew up after China adopted policies of reform and greater openness were more influenced by Western lifestyle and values and may have a different pattern of expressing emotional distress than the older generation in China. Second, we used online recruitment of the sample, which has the advantage of being economical and fast but also the disadvantage of the self-selection effect of participants. For organizational reasons, the Chinese students did not receive financial compensation for participating in the study and this could lead to bias in the data. Furthermore, the scales were found to be partially measurement invariant and to fulfill the prerequisite for comparison of latent means by including only unbiased items, which can lead to shortcomings in the interpretation of cross-cultural comparisons.

Conclusions

In summary, our findings imply that the PHQ-9 and GAD-7 could be considered as construct invariant for students across Chinese and German cultures, with individual items showing cultural differences, and thus could be used for cross-cultural comparison. The PHQ-15 did not show scalar invariance. Full scalar invariance is generally difficult to find, especially across strongly contrasting cultures. This may be due to translation problems for certain items, cultural bias in understanding certain concepts and problems with the method for testing measurement invariance. Intercultural cooperation should be encouraged in order to improve the diagnostic instruments, which are more sensitive to culturally specific symptoms. Future studies may consider alternative approaches to test measurement invariance and more research into the influence of culture on shaping somatic awareness is required. Furthermore, it is necessary to examine the universality of the scales across diverse aged populations. Previous studies demonstrated that there are qualitative differences in the symptom presentation of depression and anxiety in younger and older adults, and that the different presentations of depression and anxiety in older adults are not fully assessed by the current measures of depression and anxiety [69, 70]. Our study is one of the first to investigate the measurement invariance of the frequently used PHQ-15, PHQ-9 and GAD-7 in large groups in China and Germany, which suggests that the constructs of a subject (e.g., somatic symptoms) could vary in its expression in different cultural contexts and that measurement equivalence of the measurement instrument should be ensured in comparative cultural studies.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12888-020-02859-8.

Additional file 1: Table S1. Items of the four first-order latent factors of the PHQ-15

Abbreviations

PHQ-15: Patient Health Questionnaire-15 Physical Symptoms; PHQ-9: Patient Health Questionnaire-9 Depression Symptoms; GAD-7: Generalized Anxiety Disorder 7-item scale; CFI: Comparative fit index; RMSEA: Root mean square error of approximation; SRMS: Standardized root mean residual; ΔCFI: The difference in CFI between the base model and the constrained model

Acknowledgements

We would like to thank the students of University of Zhengzhou and University of Marburg for their participation.

Authors' contributions

YZ and WR were responsible for the overall conception, design and analysis of this study. JX contributed to sample preparation. YZ made contribution to

the revising of statistical analysis and interpretation of the data and wrote the manuscript. WR provided critical feedback and helped shape the manuscript. All authors have read and approved the manuscript

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
This study was carried out in accordance with the recommendations of "Ordnung für die Lokale Ethik-Kommission des Fachbereichs Psychologie vom 10.02.2010", die Lokale Ethik-Kommission (LEK) of University of Marburg, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the 'Ethic-Committee of Department of Psychology of University Marburg' (reference number: 2016-19 k).

Consent for publication
Not applicable.

Competing interests
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details
[1]Clinical Psychology and Psychotherapy, Department of Psychology, Philipps-University of Marburg, Gutenbergstr. 18, D-35032 Marburg, Germany. [2]Department of Marxism, University of Zhengzhou, Zhengzhou, China.

Received: 22 April 2020 Accepted: 8 September 2020
Published online: 01 October 2020

References

1. Bromet E, Andrade LH, Hwang I, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. 2011;9. https://doi.org/10.1186/1741-7015-9-90.
2. Demyttenaere K, Bruffaerts R, Posada-Villa J, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization world mental health surveys. J Am Med Assoc. 2004;291(21):2581–90. https://doi.org/10.1001/jama.291.21.2581.
3. Baxter AJ, Scott KM, Vos T, Whiteford HA. Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol Med. 2013;43(5):897–910. https://doi.org/10.1017/S003329171200147X.
4. Ferrari AJ, Somerville AJ, Baxter AJ, et al. Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. Psychol Med. 2013;43(3):471–81. https://doi.org/10.1017/S0033291712001511.
5. Hofmann SG, Asnaani A, Hinton DE. Cultural aspects in social anxiety and social anxiety disorder. Depress Anxiety. 2010;27(12):1117–27. https://doi.org/10.1002/da.20759.
6. Khoo EM, Mathers NJ, McCarthy SA, Low WY. Somatisation disorder and its associated factors in multiethnic primary care clinic attenders. Int J Behav Med. 2012;19(2):165–73. https://doi.org/10.1007/s12529-011-9164-7.
7. Liu L, Bi B, Qin X, et al. The prevalence of somatoform disorders in internal medicine outpatient departments of 23 general hospitals in Shenyang, China. Gen Hosp Psychiatry. 2012;34(4):339–44. https://doi.org/10.1016/j.genhosppsych.2012.02.002.
8. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry. 2019;6(3):211–24. https://doi.org/10.1016/S2215-0366(18)30511-X.
9. Jacobi F, Höfler M, Strehle J, et al. Twelve-month prevalence, comorbidity and correlates of mental disorders in Germany: the mental health module of the German health interview and examination survey for adults (DEGS1-MH). Int J Methods Psychiatr Res. 2014;23(3):304–19. https://doi.org/10.1002/mpr.1439.
10. Busch MA, Maske UE, Ryl L, Schlack R, Hapke U. Prevalence of depressive symptoms and diagnosed depression among adults in Germany. Results of the German health interview and examination survey for adults (DEGS1). Bundesgesundheitsblatt - Gesundheitsforsch - Gesundheitsschutz. 2013;56(5–6):733–9. https://doi.org/10.1007/s00103-013-1688-3.
11. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Washington, DC: Author; 2013.
12. World Health Organization. The ICD-10 classification of mental and Behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.
13. Simon GE, Goldberg DP, Von Korff M, Üstün TB. Understanding cross-national differences in depression prevalence. Psychol Med. 2002;32:585–94. https://doi.org/10.1017/s0033291702005457.
14. Löwe B, Spitzer RL, Zipfel S, Herzog W. PHQ-D Gesundheitsfragebogen Für Patienten: Manual Komplettversion Und Kurzform- Autorisierte Deutsche Version Des Prime MD Patient Health Questionnaire (PHQ). 2nd ed. Heidelberg: Pfizer; 2002.
15. Online Assessment Measures. https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/assessment-measures. Accessed July 18, 2020.
16. Chen S, Chiu H, Xu B, et al. Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. Int J Geriatr Psychiatry. 2010;25(11):1127–33. https://doi.org/10.1002/gps.2442.
17. Gräfe K, Zipfel S, Herzog W, Löwe B. Screening for psychiatric disorders with the patient health questionnaire (PHQ). Results from the German validation study. Diagnostica. 2004;50:171–81. https://doi.org/10.1026/0012-1924.50.4.171.
18. Lee S, Ma YL, Tsang A. Psychometric properties of the Chinese 15-item patient health questionnaire in the general population of Hong Kong. J Psychosom Res. 2011;71(2):69–73. https://doi.org/10.1016/j.jpsychores.2011.01.016.
19. Löwe B, Decker O, Müller S, et al. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med Care. 2008;46(3):266–74. https://doi.org/10.1097/MLR.0b013e318160d093.
20. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. Gen Hosp Psychiatry. 2006;28(1):71–7. https://doi.org/10.1016/j.genhosppsych.2005.07.003.
21. Tong X, An D, Mcgonigal A, Park S, Zhou D. Validation of the Generalized Anxiety Disorder-7 (GAD-7) among Chinese people with epilepsy. Epilepsy Res. 2016;120:31–6. https://doi.org/10.1016/j.eplepsyres.2015.11.019.
22. Zhang L, Fritzsche K, Liu Y, et al. Validation of the Chinese version of the PHQ-15 in a tertiary hospital. BMC Psychiatry. 2016;16(89). https://doi.org/10.1186/s12888-016-0798-5.
23. Ryder AG, Yang J, Zhu X, et al. The cultural shaping of depression: somatic symptoms in China, psychological symptoms in North America? J Abnorm Psychol. 2008;117(2):300–13. https://doi.org/10.1037/0021-843X.117.2.300.
24. Chang SM, Hahm B, Lee J, et al. Cross-national difference in the prevalence of depression caused by the diagnostic threshold. J Affect Disord. 2008;106(1–2):159–67. https://doi.org/10.1016/j.jad.2007.07.023.
25. Bieda A, Hirschfeld G, Schönfeld P, Brailovskaia J, Zhang XC, Margraf J. Universal happiness? Cross-cultural measurement invariance of scales assessing positive mental health. Psychol Assess. 2017;29(4):408–21. https://doi.org/10.1037/pas0000353.
26. Miller MJ, Sheu HB. Conceptual and measurement issues in multicultural psychology research. In: Brown SD, Lent RW, editors. Handbook of counseling psychology. 4th ed: Wiley; 2008. p. 103–20.
27. Brown TA. Confirmatory factor analysis for applied research. New York: Guilford; 2006.
28. Bryne BM, Shavelson RJ, Muthén B. Testing for the equivalence of fator covariance and mean structures : the issue of partial measurement in variance. Psychol Bull. 1989;105(3):456–66. https://doi.org/10.1037/0033-2909.105.3.456.
29. Muthén B, Christoffersson A. Simultaneous factor analysis of dichotomous variables in several groups. Psychometrika. 1981;46(4):407–19.
30. Witthöft M, Fischer S, Jasper F, Rist F. Clarifying the latent structure and correlates of somatic symptom distress: A bifactor model approach. Psychol. Assess. 2016;28(1):109–15. https://doi.org/10.1037/pas0000150.
31. Mewes R, Christ O, Rief W, Brähler E, Martin A, Glaesmer H. Are depression and somatisation equivalent for migrants and native Germans? An investigation of measurement invariance for the PHQ-9 and PHQ-15. Diagnostica. 2010;56:230–9. https://doi.org/10.1026/0012-1924/a000026.

32. Leonhart R, de Vroege L, Zhang L, et al. Comparison of the factor structure of the Patient Health Questionnaire for somatic symptoms (PHQ-15) in Germany, the Netherlands, and China. A transcultural structural equation modeling (SEM) study. Front. Psychiatry. 2018. https://doi.org/10.3389/fpsyt.2018.00240.
33. Hirsch O, Donner-Banzhoff N, Bachmann V. Measurement equivalence of four psychological questionnaires in native-born Germans, Russian-speaking immigrants, and native-born Russians. J Transcult Nurs. 2013. https://doi.org/10.1177/1043659613482003.
34. Keum BT, Miller MJ, Inkelas KK. Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse U.S. college students. Psychol Assess. 2018;30(8):1096–106. https://doi.org/10.1037/pas0000550.
35. Reich H, Rief W, Brähler E, Mewes R. Cross-cultural validation of the German and Turkish versions of the PHQ-9: an IRT approach. BMC Psychol. 2018; 6(26). https://doi.org/10.1186/s40359-018-0238-z.
36. Baas KD, Cramer AOJ, Koeter MWJ, Van De Lisdonk EH, Van Weert HC, Schene AH. Measurement invariance with respect to ethnicity of the patient health Questionnaire-9 (PHQ-9). J Affect Disord. 2011;129(1–3):229–35. https://doi.org/10.1016/j.jad.2010.08.026.
37. Galenkamp H, Stronks K, Snijder MB, Derks EM. Measurement invariance testing of the PHQ-9 in a multi-ethnic population in Europe: the HELIUS study. BMC Psychiatry. 2017;17(349). https://doi.org/10.1186/s12888-017-1506-9.
38. Tibubos AN, Beutel ME, Schulz A, et al. Is assessment of depression equivalent for migrants of different cultural backgrounds? Results from the German population-based Gutenberg health study (GHS). Depress Anxiety. 2018;35(12):1178–89. https://doi.org/10.1002/da.22831.
39. Petersen JJ, Paulitsch MA, Hartig J, Mergenthal K, Gerlach FM, Gensichen J. Factor structure and measurement invariance of the patient health Questionnaire-9 for female and male primary care patients with major depression in Germany. J Affect Disord. 2015;170:138–42. https://doi.org/10.1016/j.jad.2014.08.053.
40. Doi S, Ito M, Takebayashi Y, Muramatsu K, Horikoshi M. Factorial validity and invariance of the patient health questionnaire (PHQ)-9 among clinical and non-clinical populations. PLoS One. 2018. https://doi.org/10.1371/journal.pone.0199235.
41. Zeng Q, He Y, Liu H, et al. Reliability and validity of Chinese version of the generalized anxiety disorder 7-item (GAD-7) scale in screening anxiety disorders in outpatients from traditional Chinese internal department. Chinese Ment Heal J. 2013;27:163–8.
42. Parkerson HA, Thibodeau MA, Brandt CP, Zvolensky MJ, Asmundson GJG. Cultural-based biases of the GAD-7. J Anxiety Disord. 2015;31:38–42. https://doi.org/10.1016/j.janxdis.2015.01.005.
43. Zhou Y, Lemmer G, Xu J, Rief W. Cross-cultural measurement invariance of scales assessing stigma and attitude to seeking professional psychological help. Front Psychol. 2019. https://doi.org/10.3389/fpsyg.2019.01249.
44. Kroenke K, Spitzer R, Williams JBW. The PHQ-15 : Validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med. 2002; 64(2):258–66.
45. Kroenke K, Spitzer R, Williams W. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
46. Wang W, Bian Q, Zhao Y, et al. Reliability and validity of the Chinese version of the patient health questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry. 2014;36(5):539–44. https://doi.org/10.1016/j.genhosppsych.2014.05.021.
47. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder-the GAD-7. Arch Intern Med. 2006;166:1092–7. https://doi.org/10.1001/archinte.166.10.1092.
48. Kroenke K, Spitzer RL, Williams JBW, Monahan P, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med. 2007;146(5):317–25. https://doi.org/10.7326/0003-4819-146-5-200703060-00004.
49. He X, Li C, Qian J, Cui HS. Reliability and validity of a generalized anxiety scale in general hospital outpatients. Shanghai Arch Psychiatry. 2010;22(4): 200–3. https://doi.org/10.3969/j.issn.1002-0829.2010.04.002.
50. Muthén LK, Muthén BO. Mplus User ' s Guide. Muthén & Muthén; 2015.
51. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118.
52. Bentler PM. Comparative fit indexes in structural models. Psychol Bull. 1990; 107(2):238–46.
53. Weiber R, Mühlhaus D. Strukturgleichungsmodellierung - Eine Anwendungsorientierte Einführung in Die Kausalanalyse Mit Hilfe von AMOS, SmartPLS Und SPSS (Structural Equation Modeling - An Application-Oriented Introduction to Causal Analysis Using AMOS, SmartPLS, and SPSS). 2nd ed. Wiesbaden: Springer Gabler; 2015.
54. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ Res Methods. 2000;3(1):4–70. https://doi.org/10.1177/109442810031002.
55. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model A Multidiscip J. 2002;9(2):233–55. https://doi.org/10.1207/S15328007SEM0902.
56. Ma-Kellams C. Cross-cultural differences in somatic awareness and interoceptive accuracy: a review of the literature and directions for future research. Front Psychol. 2014. https://doi.org/10.3389/fpsyg.2014.01379.
57. Rimé B, Philippot P, Cisamolo D. Social schemata of peripheral changes in emotion. J Pers Soc Psychol. 1990;59(1):38–49. https://doi.org/10.1037/0022-3514.59.1.38.
58. Philippot P, Rimé B. The perception of bodily sensations during emotion: a cross-cultural perspective. Polish Psychol Bull. 1997;28(2):175–88.
59. Greiff S, Scherer R. Still comparing apples with oranges?: some thoughts on the principles and practices of measurement invariance testing. Eur J Psychol Assess. 2018;34:141–4. https://doi.org/10.1027/1015-5759/a000487.
60. Marsh HW, Muthén B, Asparouhov T, et al. Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. Struct Equ Model. 2009;16(3):439–76. https://doi.org/10.1080/10705510903008220.
61. Schmitt N, Kuljanin G. Measurement invariance: review of practice and implications. Hum Resour Manag Rev. 2008;18(4):210–22. https://doi.org/10.1016/j.hrmr.2008.03.003.
62. Leung FKS. Behind the high achievement of east Asian students. Educ Res Eval. 2002;8(1):87–108. https://doi.org/10.1076/edre.8.1.87.6920.
63. Hau K-T, Ho IT. Chinese students'motivation and achievement. In: Bond MH, ed. The Oxford Handbook of Chinese Psychology: Oxford University Press; 2010. p. 187–204.
64. Parker G, Cheah YC, Roy K. Do the Chinese somatize depression? A cross-cultural study. Soc Psychiatry Psychiatr Epidemiol. 2001;36:287–93. https://doi.org/10.1007/s001270170046.
65. Westover AN, Marangell LB. A cross-national relationship between sugar consumption and major depression? Depress Anxiety. 2002;16(3):118–20. https://doi.org/10.1002/da.10054.
66. Balsamo M, Saggino A. Determining a diagnostic cut-off on the Teate depression inventory. Neuropsychiatr Dis Treat. 2014;10:987–95. https://doi.org/10.2147/NDT.S55706.
67. Ruscio AM, Hallion LS, Lim CCW, et al. Cross-sectional comparison of the epidemiology of DSM-5 generalized anxiety disorder across the globe. JAMA Psychiatry. 2017;74(5):465–75. https://doi.org/10.1001/jamapsychiatry.2017.0056.Cross-sectional.
68. Zhong J, Wang ÃA, Qian M, et al. Shame, personality, and social anxiety symptoms in Chinese and American nonclinical samples: a cross-cultural study. Depress An. 2008;25(5):449–60. https://doi.org/10.1002/da.20358.
69. Balsamo M, Cataldi F, Carlucci L, Padulo C, Fairfield B. Assessment of late-life depression via self-report measures: a review. Clin Interv Aging. 2018;13: 2021–44. https://doi.org/10.2147/CIA.S178943.
70. Balsamo M, Cataldi F, Carlucci L, Fairfield B. Assessment of anxiety in older adults: a review of self-report measures. Clin Interv Aging. 2018;13:573–93. https://doi.org/10.2147/CIA.S114100.

Publisher's Note