Reliability, Validity, and Measurement Invariance of the General Anxiety Disorder Scale Among Chinese Medical University Students

Chi Zhang [1,2*], Tingting Wang [3], Ping Zeng [1,2], Minghao Zhao [4], Guifang Zhang [1,2], Shuo Zhai [5], Lingbing Meng [2,6], Yuanyuan Wang [7] and Deping Liu [2,6*]

[1] The Key Laboratory of Geriatrics, Beijing Institute of Geriatrics, Beijing Hospital, National Center of Gerontology, National Health Commission, Beijing, China, [2] Institute of Geriatrics Medicine, Chinese Academy of Medical Sciences, Beijing, China, [3] International Student Office of International Cooperation Department, Peking University Health Science Center, Beijing, China, [4] School of Basic Medicine, Peking University Health Science Center, Beijing, China, [5] Department of Education, Beijing Hospital, National Center of Gerontology, Beijing, China, [6] Department of Cardiology, Beijing Hospital, National Center of Gerontology, Beijing, China, [7] National Center for Health Professions Education Development, Peking University Health Science Center, Beijing, China

Background: Medical students are affected by high levels of general anxiety disorder. However, few studies have specifically focused on the applicability of universal anxiety screening tools in this sample. This study was aimed to evaluate the psychometric property of the 7-item Generalized Anxiety Disorder Scale (GAD-7) among Chinese medical university students.

Methods: A questionnaire survey was conducted among 1,021 medical postgraduates from six polyclinic hospitals. Internal consistency and convergent validity of the GAD-7 were evaluated. Factor analyses were used to test the construct validity of the scale. An item response theory (IRT) framework was used to estimate the parameters of each item. Multi-group confirmatory analyses and differential item function analyses were used to evaluate the measurement equivalence of the GAD-7 across age, gender, educational status, and residence.

Results: Cronbach's α coefficient was 0.93 and the intraclass correlation coefficients ranged from 0.71 to 0.87. The GAD-7 summed score was significantly correlated with measures of depression symptoms, perceived stress, sleep disorders, and life satisfaction. Parallel analysis and confirmatory factor analysis supported the one-factor structure of the GAD-7. Seven items showed appropriate discrimination and difficulty parameters. The GAD-7 showed good measurement equivalence across demographic characteristics. The total test information of the scale was 22.85, but the test information within the range of mild symptoms was relatively low.

Conclusions: The GAD-7 has good reliability, validity, and measurement invariance among Chinese medical postgraduate students, but its measurement precision for mild anxiety symptoms is insufficient.

Keywords: general anxiety disorder scale, medical students, classical test theory, item response theory, measurement invariance

INTRODUCTION

The prevalence of mental health disorders has increased considerably among medical students including postgraduates (1). These students are affected by higher levels of anxiety than students who major in other disciplines (2–5) as well as the general population (6, 7). Anxiety has garnered little attention and is often undetected or undertreated in the general population. In particular, only a small number of college students undergo timely screening (8). Generalized anxiety disorder (GAD) is the most common form of anxiety, which is characterized by excessive and persistent worry (9, 10). Studies have shown that GAD was correlated with academic performance (11), depression symptoms (12, 13), sleep problems (14), and adverse events (15).

Several systematic reviews have described high levels of general anxiety disorder among medical students in the US (16), Canada (3), Brazil (17), and China (18). Anxiety is most prevalent among medical students from the Middle East and Asian countries (19). A recent review including 10 investigation studies showed that the prevalence of anxiety among Chinese medical students is 21%, which is higher than that of students majoring in other subjects, as well as medical students from other Asian countries (20). A cross sectional study showed that 11% of postgraduate medical residents in Bangladesh had anxiety disorders (21). Medical university students are affected by various sources of stress, such as academia, employment, family, tutors, and a harsh health service environment. Although researchers are concerned about the prevalence of anxiety disorder among medical students, more attention should be paid to the early screening and a valid tool for GAD screening needs to be generally accepted in this sample. However, the literature regarding this specific population has been relatively insufficient.

The 7-item Generalized Anxiety Disorder Scale (GAD-7) (9), recommended by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (22), is a common instrument used in the screening of generalized anxiety disorders because of its simplicity and operability. The GAD-7 has been translated into different languages, including Chinese, over the last two decades (23–26). The reliability, validity, and diagnostic capability of the GAD-7 have been confirmed, but the majority of previous psychometric studies focused on clinical settings rather than general populations (12, 23, 27–29). Although the GAD-7 has been widely used for anxiety screening among medical students (30–32), few studies have systematically evaluated its measurement properties in this sample. Besides, previous studies have focused on the psychometric performance of the overall scale, but little attention has been paid to the characteristics or measurement invariance of individual items. The measurement equivalence is an important attribute of a screening instrument, as it ensures the comparability of measurement values across different subsamples. Therefore, it is necessary to evaluate the GAD-7 comprehensively with methodologies that combine classical test theory (CTT) and item response theory (IRT). The IRT framework test the probability of subjects' response according to particular models and then evaluates parameters of the measurement tool. These methods were originally designed to evaluate examination tools and are recently widely used to assess the suitability of health-related scales (33, 34).

This study was designed to evaluate the reliability, validity, and measurement invariance of the GAD-7 using a sample of medical university students. We also aimed to provide reasonable suggestions of its application in practice.

MATERIALS AND METHODS

Participants

The study participants were 1,021 full-time medical postgraduates from six polyclinic hospitals of Peking University Medical College or Peking Union Medical College. These hospitals were Beijing Hospital, the First Hospital of Peking University, the People's Hospital of Peking University, the Third Hospital of Peking University, Peking Union Medical College Hospital, and the Cancer Hospital of The Chinese Academy of Medical Sciences. In each hospital, more than 50 percent of the total postgraduate students were selected during the survey. We estimated sample size on the basis of factor analysis and item sample ratio method. As some researchers recommend, a sample of 300–1,000 in factor analysis is excellent (35) and a sample item ratio between 10 and 20 indicates sufficient (36). Respondents in the current study included 630 (61.71%) master's and 391 (38.29%) doctoral medical students.

Procedures and Ethic

A cross-sectional questionnaire survey was conducted from April to June, 2020, and the management staff of each hospital collected the data. The ethics committee of Beijing Hospital approved this study (2020BJYYEC-231-01). All respondents gave informed consent and volunteered to participate in the study. The background and purpose of the survey, as well as informed consent, were explained on the first page of the questionnaire. In order to ensure an effective recovery rate, each respondent could receive a feedback report by email after submission. A total of 1,108 questionnaires were collected and 87 invalid questionnaires (either incomplete within the allotted time, had unidentifiable information, or too many repetitive responses) were excluded. Thus, 1,021 valid questionnaires were obtained and included in the analysis.

Measurements

General anxiety disorders were measured with a Chinese GAD-7 version. The Chinese GAD-7 version was firstly translated by He and colleagues using standard translation methods in 2010 (37). In the current study, we quoted the Chinese GAD-7 version from He's study and further translated it back into English by a doctor of psychiatry, two specialists in medical education, and one English native overseas postgraduate. These cross-cultural adaptation procedures ensured the semantic equivalence between the translated version and the original. Besides, as the statements of questions in the GAD-7 are relatively concise, university students could understand the meaning easily, thus the face validity and content validity of the Chinese GAD-7 version were supported. Previous studies have demonstrated the GAD-7's appropriate screening utility in clinical samples and the general

population (23, 27, 38, 39). Its unidimensional structure has been demonstrated in many published studies employing different methodologies (27, 29), with a few exceptions (12, 40). The GAD-7 is a 7-item self-report measurement designed to screen the presence of general anxiety disorders over the previous 2 weeks (9). Items consist of seven statements about worry or somatic tension, which are rated on a four-point Likert scale as follows: 0 (not at all); 1 (several days); 2 (more than half the number of days); and 3 (nearly every day), indicating frequency levels of GAD symptoms. The GAD-7 summed score ranges from 0 to 21, with cutoff points of 5, 10, and 15 allowing researchers to classify

the anxiety as none/normal (0–4), mild (5–9), moderate (10–14), and severe (15–21) (41). However, the cutoff score for the prevalence of general anxiety disorders has not been consistent among multiple samples. The original validation study of the GAD-7 in the primary care setting, adopted 9/10 as the cutoff score (9), while the recommended cutoff scores range from 7 to 13 for different versions (23, 26, 27, 41–44). Furthermore, a small number of studies have used a 4/5 score as an optimal cutoff (45).

The Patient Health Questionnaire 9-item depression scale (PHQ-9) (46), a valid self-administered depression screening and diagnostic tool, was used to measure depression symptoms. The

TABLE 1 | Distributions of scores of 7 items in the General Anxiety Disorder Scale [n (%)].

Contents	Not at all	Several days	More than half the number of days	Nearly every day	Mean	SD
1. Feeling nervous, anxious, or on edge	134 (13.12)	718 (70.32)	128 (12.54)	41 (4.02)	1.07	0.64
2. Becoming easily annoyed or irritable	217 (21.25)	642 (62.88)	136 (13.32)	26 (2.55)	0.92	0.67
3. Feeling afraid, as if something awful might happen	366 (35.85)	509 (49.85)	117 (11.46)	29 (2.84)	0.81	0.73
4. Worrying too much about different things	370 (36.24)	485 (47.50)	133 (13.03)	33 (3.23)	0.83	0.77
5. Being so restless that it is hard to sit still	297 (29.09)	579 (56.71)	119 (11.66)	26 (2.55)	0.88	0.71
6. Not being able to stop or control worrying	321 (31.44)	547 (53.57)	127 (12.44)	26 (2.55)	0.86	0.72
7. Trouble relaxing	345 (33.79)	506 (49.56)	136 (13.32)	34 (3.33)	0.86	0.77

SD, standard deviation.

TABLE 2 | Characteristics and summed GAD-7 score of 1,021 medical students.

Characteristics		N	Proportion (%)	Mean	SD	P-value
Age	19–25	461	45.15	6.53	3.51	0.126[†]
	26–37	560	54.85	6.09	3.42	
Gender	Male	356	34.87	6.38	3.34	0.594[†]
	Female	668	65.36	6.28	3.71	
Education	Doctor	391	38.29	6.25	3.41	0.108[†]
	Master	630	61.71	6.24	3.71	
Degree	Academic	461	44.93	6.04	3.68	0.063[§]
	Professional	563	54.82	6.54	3.43	
Residence	Rural	354	34.30	6.55	3.29	0.178[†]
	City or town	670	64.86	6.19	3.68	
Income	Poverty*	527	50.97	6.87	3.63	0.002[§]
	No-poverty	497	48.02	5.95	3.48	
Major	Clinical medicine	794	76.64	6.11	3.54	0.006[§]
	Others	230	22.18	7.03	3.89	
Institutional satisfaction	Yes	887	85.29	5.98	3.61	<0.001[§]
	No	137	13.16	8.51	3.01	
Major satisfaction	Yes	837	80.33	5.97	3.57	<0.001[§]
	No	187	17.93	7.86	3.31	
Relationship with tutors	Good	867	83.05	6.04	3.54	<0.001[§]
	Bad	157	15.02	7.85	3.61	
Exercise habits	Yes	397	37.95	5.55	3.62	<0.001[§]
	No	627	59.89	6.81	3.50	

*The annual household income is <100,000 yuan.
[†]Student's t test.
[§]Wilcoxon rank-sum test.
SD, standard deviation.

Cronbach's α coefficient of PHQ-9 in this study was 0.89. The 10-item Perceived Stress Scale (PSS-10) (47) was used to measure perceived stress. The PSS-10 is one of the most frequently used self-report psychological questionnaires, which is widely used across various cultures and populations (48). It showed appropriate consistency reliability with an α coefficient of 0.91. The Athens Insomnia Scale (AIS) (49) was used to quantify the presence of insomnia among study participants. The AIS is widely used in the general population and included eight three-point Likert items. It showed appropriate consistency reliability with an α coefficient of 0.87.

Statistical Analysis

Continuous variables were described as the mean ± standard deviation (mean ± SD), and categorical variables were described as numbers with percentages [n (%)]. The Student's t test or Wilcoxon rank sum test was used to compare the differences of GAD-7 scores among different groups. Spearman correlation coefficient was used to analyze the correlation between GAD-7 score and other measured outcomes. Statistical significance was accepted at the two-sided 0.05. Internal consistency of the scale was evaluated using the Cronbach's α coefficient and Guttmann's coefficient, with an α coefficient >0.7 indicating good internal reliability (50). An exploratory factor analysis (EFA) using the principal component method was performed to explore the factor structure. Parallel analyses (PA) (51) were used to retain factors with 500 random data matrices. The retained eigenvalues should meet the K1 criterion (≥1) and greater than the average or the 95th percentile of the random samples. Factor loading >0.6 in exploratory factor analysis (EFA) is considered acceptable. Confirmatory factor analyses (CFA) with robust weighted least squares estimation were conducted in Mplus (version 7.4) in cases of violation of the multivariate normality assumption. χ^2/df, root mean square error of approximation (RMSEA), comparative fit index (CFI), and normed fit index (NFI) were used to evaluate the fitness. The model is considered to have a good fit with a χ^2/df of 5 or less, a RMSEA of 0.1, a CFI and NFI >0.90 (52).

The IRT analysis with a fitted Semejima graded response model was implemented to estimate the discrimination (a) and difficulty (b) parameters of seven items using the IRTPRO version 4.2 software. Prior to the implementation of the IRT, the unidimensionality assumption was tested using factor analyses. The local independence was confirmed using the χ^2LD statistic (53) and residuals covariance (54). A χ^2LD statistic <10 and a standardized residual covariance <0.2 between two items indicated an acceptable level of local independence (54). The item characteristic curve (ICC) was used to establish the relationship between subjects' potential trait and their responses, and the item information curve (IIC) was used to evaluate the measurement precision through the test information function (TIF). The measurement precision of a scale is sufficient when the total test information is above 16 (55). In addition, by applying the posterior estimation method, the transformation relationship between the original sum score and IRT characteristic score was established (56).

Factorial invariance of the GAD-7 across age, gender, education, and residence was tested by a multigroup confirmatory factor analysis approach, which consisted of a series of nested confirmatory steps (57). Configural invariance (free parameters), metric invariance (constraints of equivalent factor loadings), scalar invariance (further constraints of the intercepts), and strict invariance (further constraints of residual variances) models were tested across subgroups. A no-significant $_\Delta\chi^2$ ($P > 0.05$); a $_\Delta$CFI value <0.01; and a $_\Delta$RMSEA value <0.15 were used to compare the fit of nested models (58). We examined measurement invariance of item parameters using differential item functioning (DIF) methods (59). The DIF occurred when the relationship between the latent variable and item responses differed on item parameters across subgroups. The existence of the DIF suggests that the differences between groups may not be

FIGURE 1 | Scree plot of the 7-item Generalized Anxiety Disorder Scale in parallel analysis.
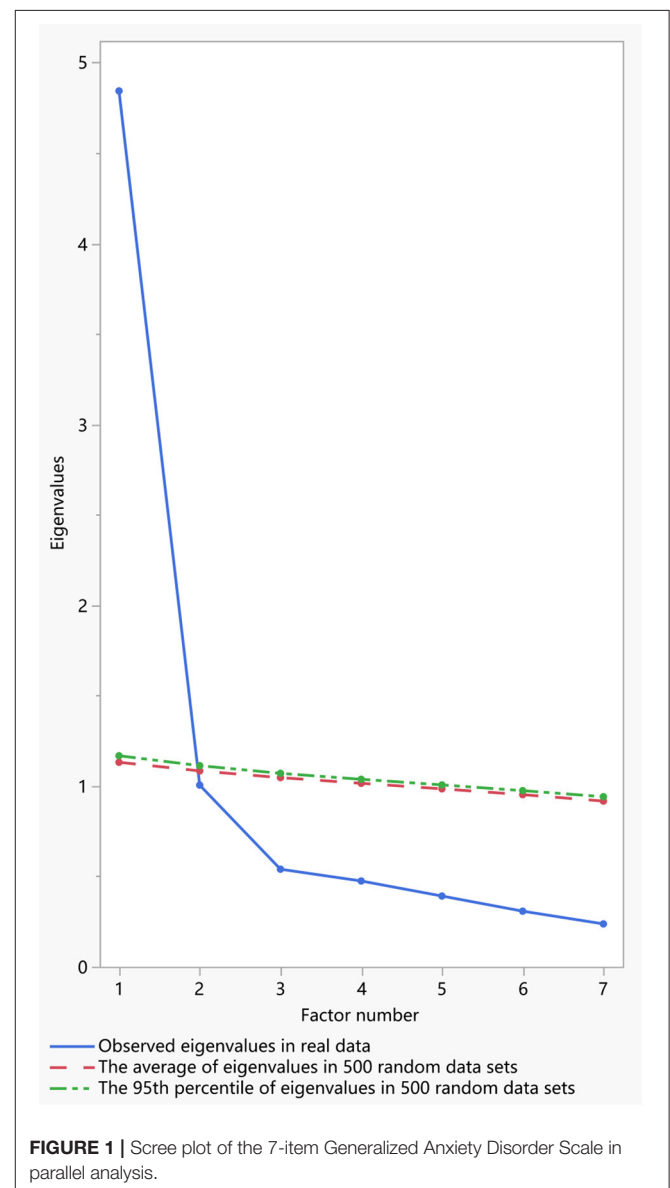
due to actual differences between groups in the survey variables, but to other factors, such as the measurement tool itself or unknown external factors (60). A no-significant $_\Delta \chi^2$ ($P > 0.05$) at specific degrees of freedom indicated acceptable parameter invariance (60).

RESULTS

Demographic Characteristics

A total of 1,021 postgraduate students (26.01 $\pm$ 2.46 years) completed the survey, including 61.71% master and 38.29% doctoral students. The majority were female (65.36%) and clinical medical students (76.64%). The average GAD-7 score was 6.29 $\pm$ 3.58 and the distribution of points for each item was described in **Table 1**. Among the participants, 34.28% (350/1,021) had no anxiety; 49.07% (501/1,021) had mild anxiety; 12.34% (126/1,021) had moderate anxiety; and 4.31% (44/1,021) had severe anxiety. The socio-demographic characteristics according to the GAD-7 scores of the postgraduate students were presented in **Table 2**.

Reliability, Validity, and Factor Structure of the GAD-7

The overall α coefficient of the GAD-7 was 0.93 and the Guttmann's coefficient was 0.89. The summed GAD-7 score was statistically significant correlated with scores of the PHQ-9 ($r = 0.78$, $P < 0.001$), PSS-10 ($r = 0.71$, $P < 0.001$), AIS-8 ($r = 0.67$, $P < 0.001$), and SWLS-5 ($r = -0.38$, $P < 0.001$). As **Table 2** showed, the α coefficient of the scale was reduced when a specific item was removed. The intraclass correlation coefficients between scores of seven items and the summed GAD-7 score ranged from 0.71–0.87 ($P < 0.001$). The KMO statistic was 0.92 and the significance of Bartlett's test of sphericity (χ^2

$= 4,997.63$, $df = 21$, $P < 0.001$) indicated that the data was suitable for factor extraction. A parallel analysis employing the principal component method was used to determine the number of factors, and one common factor was extracted. The eigenvalue of this factor was 4.82 accounting for 66.02% of the variation and the scree plot was showed in **Figure 1**. As **Table 3** showed, the loadings of seven items on this factor were >0.7. A CFA with weighted least square estimation was used to test the one-factor structure of the GAD-7. The modification index between item 3 and item 4 was 83.52, and the CFA model was modified by establishing the residual covariation correlation between the two items. The adaptability of the modified model was then significantly improved ($\chi^2/df = 3.48$, CFI = 0.97, NFI = 0.96, RMSEA = 0.05) and the factor loading of each item in the
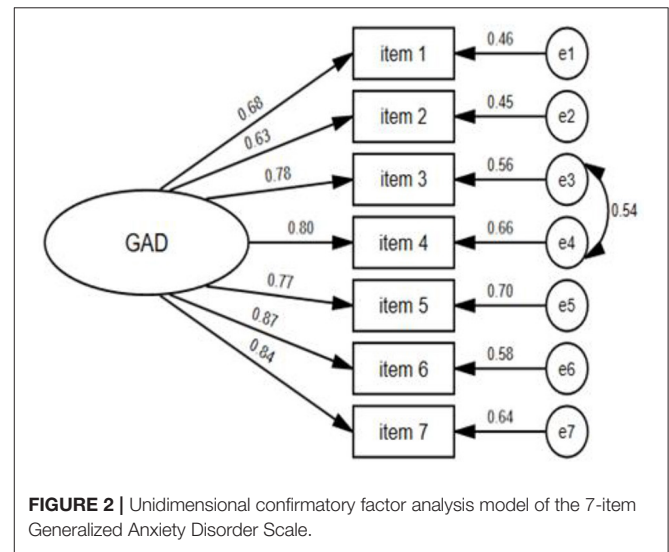
FIGURE 2 | Unidimensional confirmatory factor analysis model of the 7-item Generalized Anxiety Disorder Scale.

TABLE 3 | Item analyses of the GAD-7 based on classical test theory and item response theory.

Contents	CTT				IRT				
	α coefficient*	ICC	EFA loading	CFA loading	a (S.E.) [†]	b1 (S.E.) [‡]	b2 (S.E.) [§]	b3 (S.E.) [¶]	Information (θ)
1. Feeling nervous, anxious, or on edge	0.91	0.74	0.74	0.68	2.54 (0.26)	−1.22 (0.11)	1.69 (0.12)	2.60 (0.19)	2.49 (1.23)
2. Becoming easily annoyed or irritable	0.91	0.71	0.70	0.63	1.90 (0.17)	−0.77 (0.10)	1.87 (0.14)	3.19 (0.28)	1.16 (1.32)
3. Feeling afraid, as if something awful might happen	0.89	0.85	0.85	0.78	4.25 (0.46)	−0.18 (0.07)	1.52 (0.09)	2.58 (0.18)	3.42 (1.24)
4. Worrying too much about different things	0.89	0.86	0.86	0.80	4.79 (0.57)	−0.18 (0.07)	1.46 (0.09)	2.40 (0.15)	4.10 (2.14)
5. Being so restless that it is hard to sit still	0.90	0.79	0.79	0.77	2.68 (0.24)	−0.40 (0.08)	1.84 (0.12)	2.91 (0.23)	2.70 (1.36)
6. Not being able to stop or control worrying	0.89	0.87	0.87	0.87	4.73 (0.70)	−0.30 (0.07)	1.61 (0.09)	2.55 (0.17)	5.41 (1.15)
7. Trouble relaxing	0.90	0.84	0.84	0.84	3.51 (0.35)	−0.23 (0.07)	1.68 (0.10)	2.48 (0.17)	3.01 (1.11)

*Cronbach's α coefficients of GAD-7 when the specific item was deleted.
[†]Discrimination parameters of items.
[‡]Difficulty parameters of items for response category 1 (several days).
[§]Difficulty parameters of items for response category 2 (more than half the days).
[¶]Difficulty parameters of items for response category 3 (nearly every day).
GAD-7, 7-item Generalized Anxiety Disorder Scale; CTT, classical test theory; IRT, item response theory; ICC, intraclass correlation coefficient; EFA, exploratory factor analysis; CFA, confirmatory factor analysis; S.E., standard error.

CFA model was >0.6. This indicated that the unidimensional structure showed excellent suitability to the data. The CFA model of the GAD-7 is shown in **Figure 2**.

We then tested the factorial invariance using the multi-group confirmatory factor analysis (MGCFA) framework. The configural invariance model was used as a basic model and three restrictive models were tested step by step. As summarized in **Table 4**, the metric invariance model and scalar invariance model showed excellent fitness across age, age, gender, education, and residence ($P > 0.05$, ΔCFI < 0.01). The strict invariance model only showed acceptable fitness across residence ($P = 0.236$, ΔCFI = 0.004).
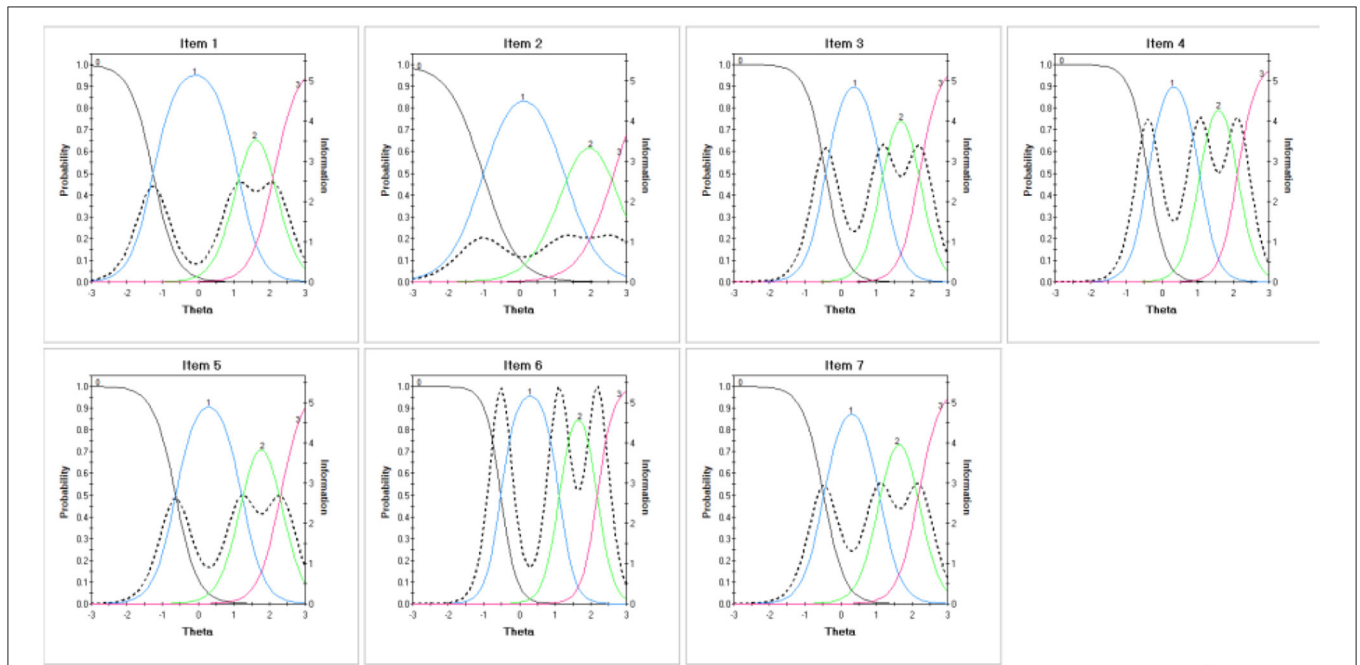
Item Characteristics of the GAD-7

As the results of EFA and CFA supported the unidimensional structure of the GAD-7, we further conducted the χ^2LD statistic matrix (**Supplementary Table 1**) and residual covariance matrix (**Supplementary Table 2**) between any two items to test its local independence. The χ^2LD statistics were all <10 (0.42 to 9.30)

and the residual covariances were <0.2 (-0.007 to 0.053), which indicated an appropriate local independence feature of the GAD-7. Among the two matrices, items 3 and 4 showed the highest χ^2LD statistic (9.3) and highest residual covariance (0.053). An item response analysis with a Semejima graded model was used to estimate the parameters of seven items. The discrimination parameter of seven items ranged from 1.90 to 4.79, and the difficulty parameter ranged from -1.22 to 3.19 with a monotonically increasing trend. All seven items had sufficient test information with a corresponding local trait (θ). We summarized the ICC and IIC of seven items in **Figure 3** and parameter values are shown in **Table 3**. We listed the conversion between the original GAD summed scores and the IRT trait scores (**Supplementary Table 3**), and divided the horizontal coordinate of the test information curve of the GAD-7 into four anxiety category levels. As **Figure 4** shows, the total test information of the GAD-7 among medical postgraduate students was 22.85, and the corresponding latent trait level located at 1.38. However, the test information within the

TABLE 4 | Factorial invariance analyses of the 7-item Generalized Anxiety Disorder Scale across age, gender, education, and residence.

Models	Metric invariance model					Scalar invariance model					Strict invariance model				
Index	$\Delta\chi^2$	df	P	ΔCFI	$_\Delta$RMSEA	$\Delta\chi^2$	df	P	ΔCFI	$_\Delta$RMSEA	$\Delta\chi^2$	df	P	ΔCFI	$_\Delta$RMSEA
Age	11.4	6	0.076	0.002	0.002	11.4	7	0.121	0.002	0.003	73.9	15	0.000	0.015	0.214
Gender	8.23	6	0.222	0.002	0.003	8.27	7	0.309	0.002	0.004	50.4	15	0.000	0.010	0.185
Education	8.44	6	0.208	0.002	0.005	11.4	7	0.120	0.002	0.003	48.0	15	0.000	0.009	0.179
Residence	5.68	6	0.460	0.001	0.004	12.5	7	0.085	0.002	0.004	18.5	15	0.236	0.004	0.005

CFI, comparative fit index; RMSEA, root mean square error of approximation.

FIGURE 3 | Item characteristic curves and item information curves of seven items in the GAD-7.

FIGURE 4 | Test information function curves of the GAD-7 according to multiple subgroups.
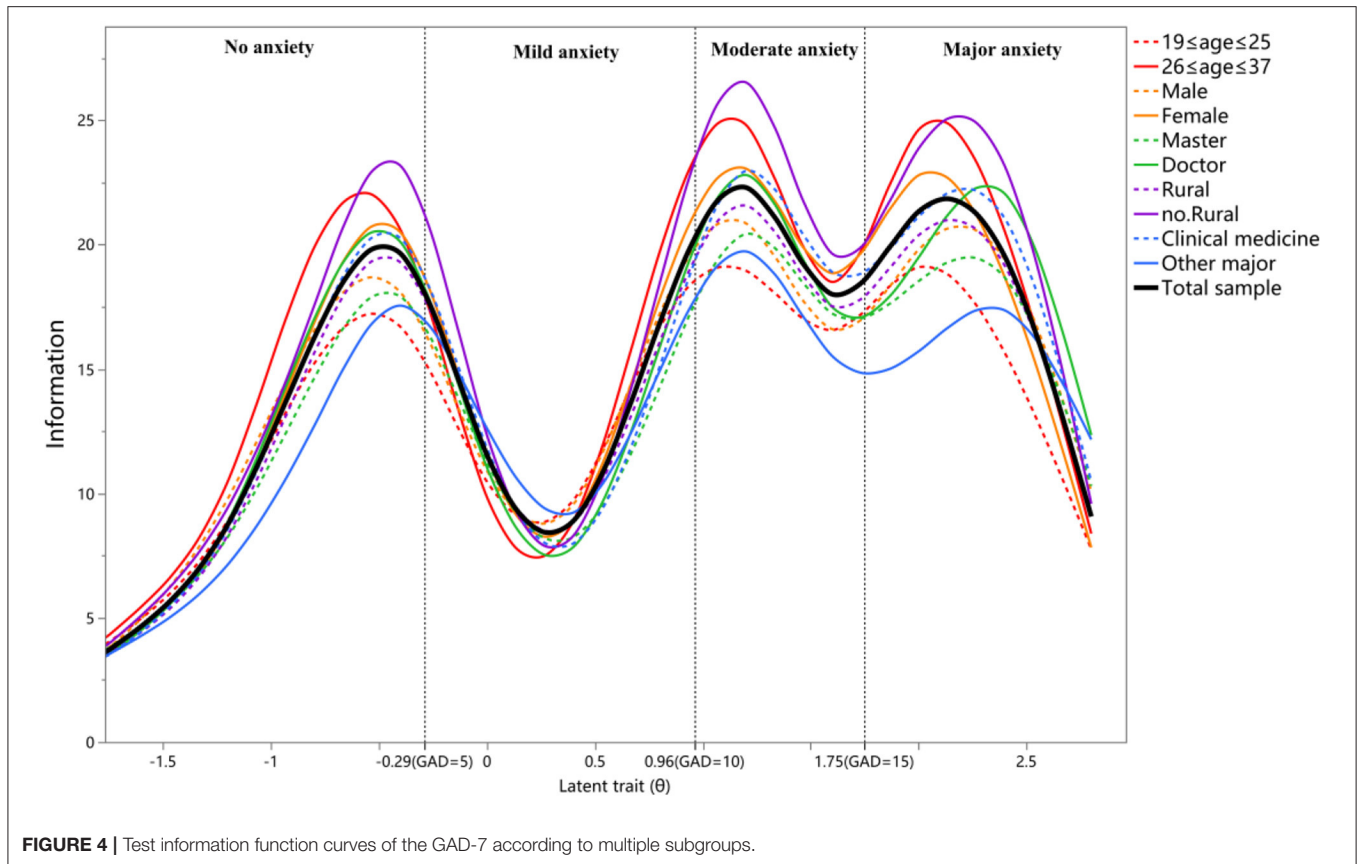
range of mild anxiety symptoms (5 ≤ GAD score <10) was relatively low.

We further analyzed the differential item function of each item across four subgroups. The results of parameter invariance are summarized in **Table 5**. No statistically significant differences were found in either discrimination or difficulty parameters ($P > 0.05$) according to the $_\triangle \chi^2$, which indicated excellent equivalence for the seven items. Furthermore, the test information function curves of the different subgroups were close to the curve of the total sample (**Figure 4**). These curvilinear paths further supported the measurement invariance of the GAD-7.

DISCUSSION

As far as the authors know, this is the first study to evaluate the psychometric properties of the GAD-7 among medical university students combining CTT and IRT. We observed a higher prevalence of general anxiety disorders (65.72%) than previous reports in China (18). This indicated that psychological impairment was a common problem among Chinese medical students. In addition, the high incidence could also be attributed to the influence of COVID-19, as teaching tasks in universities of Beijing had not fully recovered during the survey period. Notably, the differences observed in the anxiety detection rate was related to the selection of the GAD-7's cutoff (16). When the threshold of 9/10 was applied, the detection rate of general anxiety disorder dropped to 16.65%. The results of IRT analysis

showed that the GAD-7 had considerably lower test information for subjects with mild anxiety symptoms. This innovative finding supported the importance of careful selection of cutoff values in clinical practice, and the necessity of clinical diagnosis in subjects with mild symptoms (GAD scores ranging from 5 to 10). No significant differences were found in the subjects' GAD scores across different age, gender, education status, or residence subgroups. This indicated weak associations between general anxiety and demographic characteristic among the medical university students. The GAD scores were closely related to family income and satisfaction (with college, major, or tutor), and were consistent with the results of previous studies of medical students (61). The above results show that negative emotions among medical students are an important potential risk factor of general anxiety disorders.

In the current study, we implemented standardized back-translation and cross-cultural adaptation procedures to ensure the content validity of the Chinese GAD-7 version (36). The GAD-7 had a good internal consistency reliability coefficient of 0.93, which is consistent with that reported in previous studies ranging from 0.74 (42) to 0.94 (25). The strong correlation coefficient between the GAD-7 and PHQ-9 ($r = 0.78$) has also been observed among other samples (12, 13, 62). These findings suggest that anxiety disorders frequently occur alongside depression symptoms. Several previous studies have also confirmed the association between the GAD-7 and factors such as stress (12), sleep disorders (28), and life

TABLE 5 | Measurement invariance of item parameters of the GAD-7 across subgroups.

Subgroups	Item	Total$_\Delta \chi^{2*}$	df	P	$_\Delta \chi_a^{2\dagger}$	df	p	$_\Delta \chi^2 b^\S$	df	p-value
Age	1	4.2	4	0.386	1.4	1	0.233	2.7	3	0.437
	2	2.7	4	0.607	2.0	1	0.158	0.7	3	0.869
	3	4.6	4	0.327	1.3	1	0.260	3.4	3	0.339
	4	0.8	4	0.933	0.1	1	0.808	0.8	3	0.855
	5	5.3	4	0.262	2.5	1	0.134	0.8	3	0.851
	6	5.7	4	0.227	2.7	1	0.130	1.0	3	0.811
	7	2.1	4	0.724	0.7	1	0.397	1.3	3	0.719
Gender	1	4.5	4	0.338	0.8	1	0.361	3.7	3	0.296
	2	1.7	4	0.799	0.4	1	0.526	1.3	3	0.741
	3	0.4	4	0.979	0.2	1	0.648	0.2	3	0.972
	4	1.5	4	0.830	0.2	1	0.679	1.3	3	0.726
	5	2.1	4	0.716	0.0	1	0.915	2.1	3	0.552
	6	1.5	4	0.828	0.5	1	0.490	1.0	3	0.798
	7	2.8	4	0.587	0.7	1	0.414	2.2	3	0.540
Education	1	4.4	4	0.360	0.3	1	0.610	4.1	3	0.252
	2	2.8	4	0.593	0.1	1	0.726	2.7	3	0.445
	3	3.4	4	0.489	0.2	1	0.637	3.2	3	0.362
	4	4.1	4	0.391	0.0	1	0.907	4.1	3	0.251
	5	0.5	4	0.976	0.0	1	0.909	0.5	3	0.926
	6	0.5	4	0.970	0.0	1	0.958	0.5	3	0.912
	7	7.0	4	0.134	3.6	1	0.091	2.4	3	0.495
Residence	1	4.5	4	0.348	1.2	1	0.271	3.3	3	0.356
	2	3.4	4	0.497	2.1	1	0.145	1.3	3	0.741
	3	4.6	4	0.337	3.1	1	0.077	1.4	3	0.696
	4	5.8	4	0.218	2.7	1	0.101	3.1	3	0.381
	5	2.7	4	0.608	0.3	1	0.616	2.5	3	0.484
	6	1.1	4	0.896	0.8	1	0.387	0.3	3	0.953
	7	2.3	4	0.672	0.3	1	0.568	2.0	3	0.569

*$_\Delta \chi^2$ statistic for the specific item.
†$_\Delta \chi^2$ statistic for discrimination parameter(a) of the specific item.
§$_\Delta \chi^2$ statistic for difficulty parameter(b) of the specific item.

satisfaction (28). Significant correlations between the GAD-7 and theoretically related measurements support the scale's convergent validity and discrimination ability for subjects with differences in psychological status. These results are consistent with those of previous studies of multiple populations (13, 27, 41, 63). Although the one-dimensional structure of the GAD-7 proposed by its developer is not consistent across all studies, the construct validity of its unidimensional structure was confirmed in the current study. This finding was consistent with the majority of published studies conducted in both the primary care setting and the general population (9, 25, 63). Nevertheless, a two-factor structure was reported by Satomi (40) among Japanese adult populations, as well as Kertz's (12) study in an acute psychiatric sample. Heterogeneity of the sample and differences in methodology may explain these conflicting results. We modified the CFA model by establishing a residual correlation between item 3 ("Feeling afraid as if something awful might happen") and 4 ("Worrying too much about different things"). The residual pair between specific items is a common

method used to improve the scale's fitness, and has been applied in previous studies among Portuguese college students (13), American outpatients (64), and heterogeneous psychiatric populations (65). There was some overlap between content of item 3 and 4. Furthermore, the LDχ^2 statistic between the two items (9.3) was higher than that of other pairs. This indicated that these two questions reflected an ambiguous trait, other than general anxiety (e.g., fear). When we removed either of the two items, the total test information of the GAD-7 was significantly reduced. Thus, we recommend retaining all items for specific applications. The metric and scalar invariance models of the unidimensional GAD-7 showed excellent equivalence across subgroups, which has also been confirmed in various clinical and general population studies (12, 13, 39, 64). However, the strict invariance model was not equivalent across demographic characteristics. This might be related to the heterogeneity of residual covariance between different items among subgroups.

The GAD-7 has good local independence among medical university students. This is an important characteristic of an ideal

scale and one of the preconditions for IRT analysis that is often ignored by researchers (27). The difficulty and discrimination parameters of seven items were within an appropriate range (**Table 3**) and the total test information of the scale was relatively high (22.85). These findings are consistent with those of Zhong's study in pregnant women (27). Although some psychological experts suggested collapsing the response categories, "more than half the days" and "nearly every day" in a graded response scale, owing to potentially disordered thresholds (66, 67), the difficulty parameters of the four response categories increased monotonically in the present results. This indicated that response categories were used in a reasonable and ordered manner. According to the summarized ICC (**Figure 3**), the curves of the four response categories 0–4 were significantly spaced, which is inconsistent with the findings of another study of antepartum women in two low-income countries (68). The seven items showed appropriate measurements in DIF analyses (**Table 5**). This indicated that the GAD-7 was fair among the subsamples, which is consistent with the findings of Pascal's study among primary care patients (69). Besides, we used the test information curves (**Figure 4**) to describe the measurement precision of the GAD-7, according to different screening outcomes, which is helpful in choosing optimal cutoff scores. One novelty of the present study is the fact that the GAD-7 had a lower precision for persons with mild anxiety symptoms (with scores ranging from 5–10). Barthel also confirmed that the GAD-7 items measured well at higher anxiety levels, but not as well at lower levels (68). These findings strengthen the necessity for the clinical diagnosis of persons with mild anxiety symptoms and rigorous exposition of cutoff scores in practical applications. The GAD-7 had relatively sufficient test information within the range of non-anxiety symptoms, as well as moderate and major symptoms, indicating that it is a valid screening tool for the sample. We further constructed the IIC (**Figure 4**) across different demographic characteristics, and the basic shapes of all curves were relatively similar. Moreover, the test information curves among subgroups showed very little fluctuation around the curve of the total sample, which also supported its measurement equivalence.

This study had some limitations. Firstly, we did not confirm the inter-intra rater reliability and concurrent validity of the Chinese GAD-7 version. Secondly, the optimal cutoff was not identified owing to the lack of clinical diagnoses. Thirdly, the sample used in this study originated from only one city in China, and extrapolation to other populations needs to be further verified. In the future study, we will test the screening ability of specific anxiety scales in conjunction with clinical diagnosis, as well as expand the scope of random sampling nationwide.

CONCLUSION

The 7-item General Anxiety Disorder Scale showed acceptable reliability, validity, and measurement invariance among Chinese medical postgraduates. The optimal cutoff score of the GAD-7 should be considered with caution, because of its insufficient measurement precision for symptoms of mild anxiety.

DATA AVAILABILITY STATEMENT

The original contributions generated for this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Beijing Hospital approved this study (2020BJYYEC-231-01). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CZ, TW, and DL proposed the concept and design. CZ analyzed and interpreted the data, wrote the manuscript. PZ, MZ, GZ, LM, YW, and SZ drafted and edited the manuscript. CZ and DL supervised the study and obtained funding. All authors read and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.648755/full#supplementary-material

REFERENCES

1. Rotenstein LS, Ramos MA, Torre M, Segal JB, Peluso MJ, Guille C, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *JAMA*. (2016) 316:2214–36. doi: 10.1001/jama.2016.17324

2. Benton Robertson Sherry A, Tseng John M, Wen-ChihNewton Benton Fred B, Stephen L. Changes in counseling center client problems across 13 years. *Prof Psychol Res Pract*. (2003) 34:66–72. doi: 10.1037/0735-7028.34.1.66

3. Dyrbye LN, Thomas MR, Shanafelt TD. Systematic review of depression, anxiety, and other indicators of psychological distress among US, and Canadian medical students. *Acad Med*. (2006) 81:354–73. doi: 10.1097/00001888-200604000-00009

4. Stewart-Brown S, Evans J, Patterson J, Petersen S, Doll H, Balding J, et al. The health of students in institutes of higher education: an important and neglected public health problem? *J Public Health Med.* (2000) 22:492–9. doi: 10.1093/pubmed/22.4.492

5. Mehanna Z, Richa S. Prevalence of anxiety and depressive disorders in medical students. Transversal study in medical students in the Saint-Joseph University of Beirut. *Encephale.* (2006) 32:976–82. doi: 10.1016/s0013-7006(06)76276-5

6. Vontver L, Irby D, Rakestraw P, Haddock M, Prince E, Stenchever M. The effects of two methods of pelvic examination instruction on student performance and anxiety. *J Med Educ.* (1980) 55:778–85. doi: 10.1097/00001888-198009000-00007

7. Knight RG, Waal-Manning HJ, Spears GF. Some norms and reliability data for the State–Trait Anxiety Inventory and the Zung Self-Rating Depression scale. *Br J Clin Psychol.* (1983) 22:245–9. doi: 10.1111/j.2044-8260.1983.tb00610.x

8. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Bruffaerts R. Mental disorders among college students in the World Health Organization World Mental Health Surveys – CORRIGENDUM. *Psychol Med.* (2017) 3:1. doi: 10.1017/S0033291717001039

9. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* (2006) 166:1092–7. doi: 10.1001/archinte.166.10.1092

10. Do Lynna Lan Tien Nguyen: American Psychiatric Association Diagnostic Statistical Manual of Mental Disorders (DSM-IV). In: Goldstein S, Naglieri JA, editors. *Encyclopedia of Child Behavior Development.* Boston, MA: Springer US(2011). doi: 10.1007/978-0-387-79061-9_113

11. Calvo MG, Carreiras M. Selective influence of test anxiety on reading processes. *Br J Psychol.* (1993) 84:375–88. doi: 10.1111/j.2044-8295.1993.tb02489.x

12. Kertz S, Bigda-Peyton J, Bjorgvinsson T. Validity of the Generalized Anxiety Disorder-7 scale in an acute psychiatric sample. *Clin Psychol Psychother.* (2013) 20:456–64. doi: 10.1002/cpp.1802

13. Bártolo A, Monteiro S, Pereira A. Factor structure and construct validity of the Generalized Anxiety Disorder 7-item (GAD-7) among Portuguese college students. *Cad Saude Publica.* (2017) 33:e00212716. doi: 10.1590/0102-311x00212716

14. Boehm MA, Lei QM, Lloyd RM, Prichard JR. Depression, anxiety, and tobacco use: Overlapping impediments to sleep in a national sample of college students. *J Am Coll Health.* (2016) 64:565–74. doi: 10.1080/07448481.2016.1205073

15. Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, et al. Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry.* (2004) 61:807–16. doi: 10.1001/archpsyc.61.8.807

16. Hope V, Henderson M. Medical student depression, anxiety and distress outside North America: a systematic review. *Med Educ.* (2014) 48:963–79. doi: 10.1111/medu.12512

17. Pacheco JP, Giacomin HT, Tam WW, Ribeiro TB, Arab C, Bezerra IM, et al. Mental health problems among medical students in Brazil: a systematic review and meta-analysis. *Braz J Psychiatry.* (2017) 39:369–78. doi: 10.1590/1516-4446-2017-2223

18. Zeng W, Chen R, Wang X, Zhang Q, Deng W. Prevalence of mental health problems among medical students in China: a meta-analysis. *Medicine.* (2019) 98:e15337. doi: 10.1097/md.0000000000015337

19. Quek TT, Tam WW, Tran BX, Zhang M, Zhang Z, Ho CS, et al. The global prevalence of anxiety among medical students: a meta-analysis. *Int J Environ Res Public Health.* (2019) 16:2735. doi: 10.3390/ijerph16152735

20. Cuttilan AN, Sayampanathan AA, Ho RC. Mental health issues amongst medical students in Asia: a systematic review [2000-2015]. *Ann Transl Med.* (2016) 4:72. doi: 10.3978/j.issn.2305-5839.2016.02.07

21. Sadiq MS, Morshed NM, Rahman W, Chowdhury NF, Arafat S, Mullick MSI. Depression, anxiety, stress among postgraduate medical residents: a cross sectional observation in Bangladesh. *Iran J Psychiatry.* (2019) 14:192–7.

22. Samuel B. Diagnostic and statistical manual of mental disorders, 4th ed. (DSM-IV). *Am J Psychiatry.* (1995) 152:1228.

23. Tong X, An D, McGonigal A, Park SP, Zhou D. Validation of the Generalized Anxiety Disorder-7 (GAD-7) among Chinese people with epilepsy. *Epilepsy Res.* (2016) 120:31–6. doi: 10.1016/j.eplepsyres.2015.11.019

24. Moreno AL, DeSousa DA, Souza Ana Maria Frota Lisba P, Manfro GG, Crippa José AdS. Factor structure, reliability, and item parameters of the Brazilian-Portuguese Version of the GAD-7 Questionnaire. *Temas em Psicologia.* (2016) 24:367–76. doi: 10.9788/TP2016.1-25

25. García-Campayo J, Zamorano E, Ruiz MA, Pardo A, Pérez-Páramo M, López-Gómez V, et al. Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool. *Health Qual Life Outcomes.* (2010) 8:8. doi: 10.1186/1477-7525-8-8

26. Donker T, van Straten A, Marks I, Cuijpers P. Quick and easy self-rating of Generalized Anxiety Disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. *Psychiatry Res.* (2011) 188:58–64. doi: 10.1016/j.psychres.2011.01.016

27. Zhong QY, Gelaye B, Zaslavsky AM, Fann JR, Rondon MB, Sánchez SE, et al. Diagnostic Validity of the Generalized Anxiety Disorder - 7 (GAD-7) among pregnant women. *PLoS ONE.* (2015) 10:e0125096. doi: 10.1371/journal.pone.0125096

28. Hinz A, Klein AM, Brähler E, Glaesmer H, Luck T, Riedel-Heller SG, et al. Psychometric evaluation of the Generalized Anxiety Disorder Screener GAD-7, based on a large German general population sample. *J Affect Disord.* (2017) 210:338–44. doi: 10.1016/j.jad.2016.12.012

29. Lee B, Kim YE. The psychometric properties of the Generalized Anxiety Disorder scale (GAD-7) among Korean university students. *Psychiatry Clin Psychopharmacol.* (2019) 29:864–871. doi: 10.1080/24750573.2019.1691320

30. Milić J, Škrlec I, Milić Vranješ I, Podgornjak M, Heffer M. High levels of depression and anxiety among Croatian medical and nursing students and the correlation between subjective happiness and personality traits. *Int Rev Psychiatry.* (2019) 31:653–60. doi: 10.1080/09540261.2019.1594647

31. Moir F, Henning M, Hassed C, Moyes SA, Elley CR. A peer-support and mindfulness program to improve the mental health of medical students. *Teach Learn Med.* (2016) 28:293–302. doi: 10.1080/10401334.2016.1153475

32. AlShamlan NA, AlOmar RS, Al Shammari MA, AlShamlan RA, AlShamlan AA, Sebiany AM. Anxiety and its association with preparation for future specialty: a cross-sectional study among medical students, Saudi Arabia. *J Multidiscip Healthc.* (2020) 13:581–91. doi: 10.2147/jmdh.s259905

33. Chakravarty EF, Bjorner JB, Fries JF. Improving patient reported outcomes using item response theory and computerized adaptive testing. *J Rheumatol.* (2007) 34:1426–31.

34. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care.* (2000) 38:28–42. doi: 10.1097/00005650-200009002-00007

35. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes.* (2014) 12:176. doi: 10.1186/s12955-014-0176-2

36. Arafat Smy Chowdhury HR, Qusar Mmas Hafez MA. Cross cultural adaptation and psychometric validation of research instruments: a methodological review. *J Behav Health.* (2016) 5:129–36. doi: 10.5455/jbh.20160615121755

37. He XY, Li CB, Qian J, Cui HS, Wu WY. Reliability and validity of a generalized anxiety disorder scale in general hospital outpatients. *Shang Arch Psychiatry.* (2010) 22:200–3. doi: 10.3969/j.issn.1002-0829.2010.04.002

38. Kujanpää T, Ylisaukko-Oja T, Jokelainen J, Hirsikangas S, Kanste O, Kyngäs H, et al. Prevalence of anxiety disorders among Finnish primary care high utilizers and validation of Finnish translation of GAD-7 and GAD-2 screening tools. *Scand J Prim Health Care.* (2014) 32:78–83. doi: 10.3109/02813432.2014.920597

39. Teymoori A, Real R, Gorbunova A, Haghish EF, Andelic N, Wilson L, et al. Measurement invariance of assessments of depression (PHQ-9) and anxiety (GAD-7) across sex, strata and linguistic backgrounds in a European-wide sample of patients after Traumatic Brain Injury. *J Affect Disord.* (2020) 262:278–85. doi: 10.1016/j.jad.2019.10.035

40. Doi S, Ito M, Takebayashi Y, Muramatsu K, Horikoshi M. Factorial validity and invariance of the 7-Item Generalized Anxiety Disorder Scale (GAD-7) among populations with and without self-reported psychiatric diagnostic status. *Front Psychol.* (2018) 9:1741. doi: 10.3389/fpsyg.2018.01741

41. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment,

comorbidity, and detection. *Ann Intern Med.* (2007) 146:317–25. doi: 10.7326/0003-4819-146-5-200703060-00004

42. Sidik SM, Arroll B, Goodyear-Smith F. Validation of the GAD-7 (Malay version) among women attending a primary care clinic in Malaysia. *J Prim Health Care.* (2012) 4:5–11. doi: 10.1071/hc12005

43. Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, et al. Brief case finding tools for anxiety disorders: validation of GAD-7 and GAD-2 in addictions treatment. *Drug Alcohol Depend.* (2012) 125:37–42. doi: 10.1016/j.drugalcdep.2012.03.011

44. Simpson W, Glazer M, Michalski N, Steiner M, Frey BN. Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. *Can J Psychiatry.* (2014) 59:434–40. doi: 10.1177/070674371405900806

45. Seo JG, Park SP. Validation of the Generalized Anxiety Disorder-7 (GAD-7) and GAD-2 in patients with migraine. *J Headache Pain.* (2015) 16:97. doi: 10.1186/s10194-015-0583-8

46. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* (2001) 16:606–13. doi: 10.1046/j.1525-1497.2001.016009606.x

47. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav.* (1983) 24:385–96.

48. Reis D, Lehr D, Heber E, Ebert DD. The German Version of the Perceived Stress Scale (PSS-10): evaluation of dimensionality, validity, and measurement invariance with exploratory and confirmatory bifactor modeling. *Assessment.* (2019) 26:1246–59. doi: 10.1177/1073191117715731

49. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. *J Psychosom Res.* (2000) 48:555–60. doi: 10.1016/s0022-3999(00)00095-7

50. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res.* (1995) 4:293–307. doi: 10.1007/bf01593882

51. Price Barbara Comrey Andrew L, Lee Howard B. A first course in factor analysis. *Technometrics.* (1993) 35:453.

52. Bollen KA. *Structural Equations With Latent Variable.* New York, NY: Wiley (1989).

53. Orlando M Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. *Appl Psychol Meas.* (2000) 24:50–64.

54. Reeve Bryce B, Hays Ron D, Bjorner Jakob B, Cook Karon F, Crane Paul K, Teresi Jeanne A, et al. Psychometric evaluation and calibration of health-related quality of life item banks. *Med Care.* (2007) 45:22–31. doi: 10.1097/01.mlr.0000250483.85507.04

55. Lai JS, Cella D, Chang CH, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Qual Life Res.* (2003) 12:485–501. doi: 10.1023/a:1025014509626

56. Thomas ML. The value of item response theory in clinical assessment: a review. *Assessment.* (2011) 18:291–307. doi: 10.1177/1073191110374797

57. Samuel DB, South SC, Griffin SA. Factorial invariance of the Five-Factor Model Rating Form across gender. *Assessment.* (2015) 22:65–75. doi: 10.1177/1073191114536772

58. Cheung Gordon W, Rensvold Roger B. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Model A Multidiplinary J.* (2002) 9:233–55. doi: 10.1207/S15328007SEM0902_5

59. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev Rev.* (2016) 41:71–90. doi: 10.1016/j.dr.2016.06.004

60. Fikis DRJ, Oshima TC. Effect of purification procedures on DIF analysis in IRTPRO. *Educ Psychol Meas.* (2017) 77:415–28. doi: 10.1177/0013164416645844

61. Karaoglu N, Seker M. Anxiety and depression in medical students related to desire for and expectations from a medical career. *West Indian Med J.* (2010) 59:196–202.

62. Sousa TV, Viveiros V, Chai MV, Vicente FL, Jesus G, Carnot MJ, et al. Reliability and validity of the Portuguese version of the Generalized Anxiety Disorder (GAD-7) scale. *Health Qual Life Outcomes.* (2015) 13:50. doi: 10.1186/s12955-015-0244-2

63. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Med Care.* (2008) 46:266–74. doi: 10.1097/MLR.0b013e318160d093

64. Rutter LA, Brown TA. Psychometric properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in outpatients with anxiety and mood disorders. *J Psychopathol Behav Assess.* (2017) 39:140–6. doi: 10.1007/s10862-016-9571-9

65. Beard C, Björgvinsson T. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *J Anxiety Disord.* (2014) 28:547–52. doi: 10.1016/j.janxdis.2014.06.002

66. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum.* (2007) 57:1358–62. doi: 10.1002/art.23108

67. Forkmann T, Gauggel S, Spangenberg L, Brähler E, Glaesmer H. Dimensional assessment of depressive severity in the elderly general population: psychometric evaluation of the PHQ-9 using Rasch Analysis. *J Affect Disord.* (2013) 148:323–30. doi: 10.1016/j.jad.2012.12.019

68. Barthel D, Barkmann C, Ehrhardt S, Bindt C. Psychometric properties of the 7-item Generalized Anxiety Disorder scale in antepartum women from Ghana and Côte d'Ivoire. *J Affect Disord.* (2014) 169:203–11. doi: 10.1016/j.jad.2014.08.004

69. Jordan P, Shedden-Mora MC, Löwe B. Psychometric analysis of the Generalized Anxiety Disorder scale (GAD-7) in primary care using modern item response theory. *PLoS ONE.* (2017) 12:e0182162. doi: 10.1371/journal.pone.0182162