Factor structure and measurement invariance of the 8-item CES-D: a national longitudinal sample of Chinese adolescents

Shuxia Liu[1,2], Yuan Fang[1], Zhongyan Su[1,2], Jimin Cai[1,2] and Zhiyan Chen[1*]

Abstract

Background The 8-item Center for Epidemiologic Studies Depression Scale (CES-D 8) has been widely used to measure depressive symptoms in many large-scale surveys. Due to its brevity, it can lower costs, relieve respondent burdens, and ensure data quality. However, its factor structure and measurement invariance across gender and time among adolescents have not been adequately evaluated. This study investigated its factor structure and measurement invariance across gender and time among adolescents.

Methods The data was drawn from the China Family Panel Studies (CFPS) conducted in 2018 and 2020, with 3099 participants (46.82% girls) aged 11 to 18 in 2018. First, exploratory and confirmatory factor analyses were used to examine the factor structure of the CES-D 8. Next, multi-group confirmatory factor analysis was conducted to test its measurement invariance across gender and time. Finally, a longitudinal cross-gender test was conducted to further confirm the stability of the scale.

Results A two-factor structure was identified among the adolescents, including Negative Symptoms and Diminished Happiness Feeling. Measurement invariance across gender and time, as well as the longitudinal cross-gender invariance, was supported, with configural, factor loadings, thresholds and residual invariance.

Conclusions The factor structure of the CES-D 8 remains stable across gender and time among adolescents, indicating that it is a promising instrument for measuring depressive symptoms, especially in large-scale and longitudinal surveys.

Keywords Depression, Adolescents, Measurement invariance, CES-D 8, Longitudinal study

Background

Depression is one of the leading causes of disability and death around the world, contributing greatly to the global health-related burden [1–3]. Furthermore, there is an increasing trend in the prevalence of depression globally in recent years [4], with the increase speed among adolescents exceeding that among adults [5, 6]. Given the fact that adolescents are experiencing dramatic developments in many areas such as social relationships, emotion and cognition, adolescent depression might result in subsequent detrimental outcomes. Therefore, it is significant to identify adolescents with relatively high depressive

*Correspondence:
Zhiyan Chen
chenzy@psych.ac.cn
[1] CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, China
[2] Department of Psychology, University of Chinese Academy of Sciences, Beijing 101408, China

Liu *et al. BMC Psychiatry*   (2023) 23:868

Page 2 of 12

symptoms, thus to provide further diagnosis and targeted interventions as early as possible [7]. The present study focused on a screening instrument and tried to confirm its factor structure and stability across gender and time among adolescents, attempting to contribute to the universal screening of depressive symptoms.

Depression in adolescents

Before puberty, depression is rare, but its prevalence increases rapidly from childhood to adolescence, especially in girls [8]. Previous literature has suggested that it is highly associated with adverse developmental outcomes in later life, including (1) approximately 50% of adolescents with depression suffering depression or anxiety disorders in adulthood [9]; (2) higher risk of other mental illness or risk/criminal behaviors [10, 11]; (3) elevated probability of poor physical health and incompetency as adults [9].

Nevertheless, the long-term outcomes differ among individuals. For instance, a study revealed that the association between adolescent depression and poor mental health in adulthood might depend on the persistence or severity of the symptoms during adolescence [12]. Another prospective research found that adult psychiatric and functional outcomes were associated with cumulative exposure to depression, including the number of episodes and the average degree [13]. In addition, receiving community care or professional mental health services appeared to improve outcomes in later life [10]. Considering the above findings, it is urgent to screen and identify depressive symptoms among adolescents in order to provide effective and timely interventions [14, 15].

Measurements of depressive symptoms

Many scales have been used to screen and discern depressive symptoms in large populations. According to previous literature, there is no evidence that one measure is better than the others, and the choice may depend on numerous considerations [7, 16, 17]. For instance, Beck Depression Inventory (BDI) may be a more accurate measure of mild or "neurotic" depressions [16, 18], Patient Health Questionnaire-9 (PHQ-9) may be used to diagnose depressive symptoms and evaluate their severity [17], while the Center for Epidemiologic Studies Depression Scale (CES-D) is appropriate for measuring depressive symptoms in the general population [19]. Considering the purpose of screening depressive symptoms among general adolescents, this study focused on the CES-D.

The original CES-D consisted of 20 items, measuring four factors including "depressed affect", "positive affect", "somatic and retarded activity" and "interpersonal"

[19]. The scale was widely used in the Chinese context and showed adequate reliability and validity [20, 21]. However, it was time-consuming and burdensome for respondents in large-scale social surveys, so a short 8-item CES-D (CES-D 8) [22] was proposed to suit such surveys. It had been adopted in many large-scale surveys, such as Asset and Health Dynamics Among the Oldest Old (AHEAD) [23], Health and Retirement Study (HRS) [24] and European Social Survey (ESS) [25].

Although the CES-D 8 has been widely used, different factor models have been identified in previous literature. For instance, a two-factor model was supported in American samples, including "depressed mood" and "somatic complaints" [22, 24]. A different two-factor model was found in South Africa residents, with negative and positive items loading on "negative affect" and "diminished positive affect" respectively [26]. Additionally, a one-factor model with correlated uniqueness between two positively worded items was revealed using samples of Europeans [25, 27]. Therefore, the dimensions of the CES-D 8 appeared to be associated with cultural differences.

Moreover, most of the research above involved general population or aged adults. The latter might have energy difficulty in completing a time-consuming survey [25]. Adolescents, however, susceptible to reduced sustained attention due to "decreased motor control and increased impulsivity" [28], will also benefit from an effective and efficient instrument. Nevertheless, there have been few such studies focusing on adolescents. Together, further research is required to examine the factor structure of the CES-D 8 among adolescents in different countries. Therefore, the first goal of our study is to examine the factor structure of the CES-D 8 among adolescents based on a Chinese sample.

Gender differences in adolescent depression

Gender differences in depression (i.e., females are more likely to suffer major depressive disorder than males) might be one of the most robust conclusions in psychopathology studies [29]. It emerged from puberty and peaked at the age of 15 to 18 [30, 31]. Although many research suggested there were still gender differences when adolescents entered their young adulthood [31, 32], some studies revealed that the differences were becoming narrow and even disappeared during the developmental period [33, 34]. The mixed results implied more studies required to further explore the gender differences.

Most research has focused on the association between gender and depressive symptoms, ignoring the fact that assessments of depression per se can introduce bias. That is, measurement bias between gender may influence the examination of difference and its magnitude [35].

A study of gender differences in depression found that eliminating measurement bias sometimes resulted in different conclusions [36]. Therefore, it is necessary to test measurement invariance for an effective comparison of depression across gender. If the comparison is conducted based on latent means, scalar measurement invariance is required, otherwise, the difference across group may reflect the systematic response bias [35]. Similarly, if the comparison is conducted based on manifest means, strict measurement invariance is required, otherwise, the credibility of interpretations of the results will be undermined [37].

Targeting the CES-D 8, the examination of measurement invariance across gender showed inconsistency in previous literature. Some studies revealed strict measurement invariance across gender [25, 38], while others found partial measurement invariance [26, 36]. Furthermore, the samples in these studies were all adults, lack of empirical investigations into adolescents. To conclude, it is worthwhile to conduct a measurement invariance test across gender among adolescents. Therefore, our second goal is to examine the measurement invariance across gender among adolescents.

Depression development in adolescence
As mentioned previously, the rapid rise in the prevalence of depression occurs in adolescence, with subsequent development leading to different outcomes; therefore, it is essential to conduct longitudinal analyses to better understand its development over time. Researchers have been devoted to exploring the different developmental trajectories of depressive symptoms and the associated risk and protective factors from early adolescence to young adulthood, in order to develop more targeted strategies for prevention and intervention [39, 40].

However, the same instrument may measure different constructs of depressive symptoms at different time points throughout adolescence, as adolescents are experiencing dramatic changes in thinking modes, self-conception, social cognition and interpersonal relationships, which can affect how they feel and report their depressed mood [41]. Nonetheless, measurement invariance over time was seldom mentioned in the previous longitudinal studies. Longitudinal analyses without measurement invariance examination are not tenable, for it is unable to judge whether observed changes are caused by the development of the construct of interest or measurement bias. Therefore, it is essential to examine measurement invariance over time prior to longitudinal analyses.

Specifically, few studies have examined the measurement invariance of the CES-D 8 over time, as well as the longitudinal cross-gender invariance (i.e., both cross-gender and longitudinal measurement invariance are simultaneously tested) [42]. Without such examinations prior to longitudinal analyses, the results may not hold themselves and bias subsequent meta-analyses. Thus, our third goal is to assess measurement invariance of the CES-D 8 over time among adolescents, followed by a longitudinal cross-gender measurement invariance test.

The present study
This study focuses on the validity examination of the CES-D 8, investigating its factor structure and measurement invariance across gender and time, aiming to (1) confirm whether it is suitable for adolescents; (2) provide empirical evidence to discriminate the true different depressive symptoms from just measurement bias caused by gender and/or time [35]. Furthermore, the application of the brief scale will contribute to (1) easier and more efficient survey in a large scale to screen out the adolescents with relatively high depressive symptoms at a lower cost; (2) relieving the adolescents' respondent burden and optimizing their motivation to ensure the data quality [43].

To achieve the above goals, this study was performed as follows. First, considering the absence of related research on the factor structure of the CES-D 8 among adolescents, the factor structure was identified using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Second, the obtained factor structure was used to test measurement invariance across gender. Third, the measurement invariance over time and the longitudinal cross-gender invariance were examined to provide evidence for longitudinal studies.

Methods
Participants and procedure
The participants came from the China Family Panel Studies (CFPS), which was a nationally representative social survey, conducted by the Institute of Social Science Survey (ISSS) of Peking University [44]. The CFPS employed a multi-stage probability sampling method, extracted by the means of implicit stratification, including information on levels of community, family and individual [44]. For the general-purpose, the data was collected on a household basis, covering 94.5% of the population, who were from 25 provinces (or their administrative equivalents) in Chinese mainland [44].

The survey was conducted every two years since 2010. This research chose the data in 2018 (T1) and 2020 (T2), with the participants aged between 11 and 18 in 2018. There were 3315 adolescents taking part in the survey at T1, and 216 of them did not respond to any of the items in the instrument (reported in the next section). After

Liu *et al. BMC Psychiatry*   (2023) 23:868

Page 4 of 12

removing their data, 3099 adolescents (46.82% girls) remained, whose average age was 14.31 ($SD = 2.28$). 2663 (85.93%) of these adolescents were Han nationality, 423 (13.65%) were non-Han nationality, and 13 (0.42%) did not report their nationality. 1768 (57.05%) adolescents lived in rural areas, 1307 (42.17%) lived in urban areas, and 24 (0.77%) did not report the residence. The average family income (log transformed) was 4.74 (0.41). At T2, there were 1978 adolescents (48.08% girls) filled out the questionnaire, and their average age was 16.36 ($SD = 2.30$). Among them, 1711 (86.50%) adolescents were Han nationality, 263 (13.30%) were non-Han nationality, and 4 (0.20%) did not report their nationality. 1141 (57.84%) adolescents lived in rural areas, 816 (41.25%) lived in urban areas, and 18 (0.91%) did not report the residence. The average family income (log transformed) was 4.74 (0.42).[1]

Attrition analyses indicated that the participants who retained or dropped out at T2 did not differ significantly in gender ($\chi^2(1) = 3.47$, $p = 0.062$), age ($t(3097) = 0.27$, $p = 0.789$), nationality ($\chi^2(1) = 0.68$, $p = 0.409$), residence ($\chi^2(1) = 1.68$, $p = 0.195$) and family income ($t(3068) = -0.04$, $p = 0.971$).

In this study, the survey at T1 was launched in June 2018 and completed in May of the following year. The survey was conducted by telephone or face-to-to face conversations using computer-assisted personal interviews. Among the participants, 2606 adolescents (84.09%) were interviewed face-to-face. The survey at T2 was conducted in the same way between September and December in 2020; nevertheless, only 240 participants (12.13%) were interviewed on the spot because of the COVID-19 pandemic. All the adolescents responded by themselves.

Measures

Depressive symptoms

The CES-D 8 was used to measure depressive symptoms [22]. The participants were asked how often they experienced some mental state in the past week, with a 4-point rating scale, ranging from 1 (Never, less than one day) to 4 (most of the time, 5–7 days). 2 of the 8 items (i.e., "feel happy" and "have a happy life") were reverse-coded prior to data analysis. The internal consistency reliability of the instrument was assessed by omega (ω) coefficient, with the results $\omega_{T1} = 0.71$, $\omega_{T2} = 0.76$.

---

[1] Except family annual income, the other demographic covariates were drawn from the cross-year library of the individual core variables released with the data at T1. The variables in the cross-year library were considered more reliable than ones from a single survey, for the data cleaning personnel compared them from previous surveys and selected the most reasonable values according to certain principles [45].

Data analysis

Four steps were used to investigate the factor structure and measurement invariance of the CES-D 8 among the adolescents. First, EFA was performed on the data at T1 to identify the factor structure of the CES-D 8. Exploratory structural equation modeling (ESEM) was used because it could handle EFA with correlated residuals [46, 47]. Parallel analysis indicated that two factors should be retained (See Table S1 in the online supplements), so four models were examined, including one to two-factor ESEM models, with and without accounting for correlated residual between two reverse-coded items. Of particular note was that the one-factor model with correlated residual was Karim's [25] and Van de Velde's [27] model. Oblique rotation with GEOMIN strategy was used to obtain the ultimate factor loadings.

Second, using the data at T2, CFAs were performed to compare the factor structure found by the EFA with several competing models to select the final factor structure. The competing models included (1) Turvey's [22] and Steffick's [24] two-factor model (Model 1 in Fig. 1), (2) three correlated trait-correlated method (CTCM) models (Model 2–4 in Fig. 1), using latent method factor(s) to represent wording effects [48–51]. Specifically, Model 1 to Model 4 were also examined using the data at T1 (see Tables S2 and S3 in the online supplements).

Third, following the guidelines by Millsap and Yun-Tein [52], the measurement invariance across gender was tested using the multi-group CFA (MG-CFA) based on the best fitting model from the previous CFAs. A series of models were examined, including (1) a configural invariance model where each factor was constrained to have the same indicators across groups; (2) a metric invariance (weak invariance) model where the factor loadings were constrained to be equal across groups; (3) a threshold invariance model where the thresholds of each indicator were constrained to be equal across groups, which paralleled the scalar invariance (strong invariance) model for continuous indicators (here the items were considered as ordinal indicators); (4) a residual invariance (strict invariance) model additionally constraining equivalent residual variance across groups. These four models were hierarchical and the adjacent pairs were statistically compared to examine the measurement invariance.

Fourth, the measurement invariance over time and longitudinal cross-gender invariance was examined. The former used the guidelines by Liu et al. [53], while the latter used the joint guidelines by Millsap and Yun-Tein [52] and Liu et al. [53], referring to [42, 54] meanwhile. The details of the longitudinal cross-gender test were as follows. In the baseline model, (1) the test was performed among four groups (2 genders × 2 occasions), one of which was set up as reference group; (2)
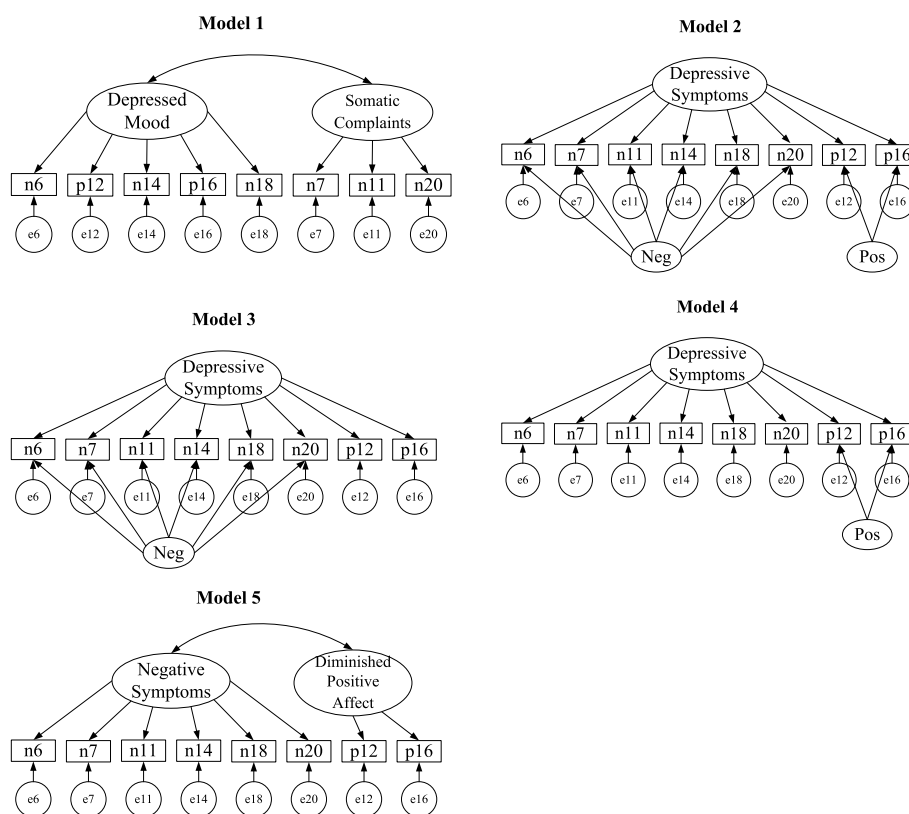
Fig. 1 Five structural equation models of CES-D 8 for the data at T2. Model 1 = Steffick's and Turvey's two-factor model; Model 2 = bi-factor model with one general factor and two specific factors measuring positive and negative method effects respectively; Model 3 = bi-factor model with one general factor and one specific factor measuring negative method effect; Model 4 = bi-factor model with one general factor and one specific factor measuring positive method effect; Model 5 = the two-factor model uncovered by the EFA

the factor loadings were freely estimated, except that the factor loading of the first indicator of each factor was set to 1; (3) the thresholds were freely estimated, except that the subsets which were constrained to be invariant across groups (i.e., one threshold for each item and a second threshold for the marker variable) [53]; (4) the factor means were freely estimated, except that the factor means were constrained to 0 in the reference group; (5) the residual variances were freely estimated, except that the residual variances were constrained to 1 in the reference group; (6) the residuals of the same items were not associated across gender groups, but associated at different time points. The constraints of the other longitudinal cross-gender invariance models could be imposed referring to the previous paragraph.

The EFA, CFAs and MG-CFAs were conducted with Mplus 7.4 [47], except that parallel analysis were conducted by R (version 4.2.1), using the package *psych* (version 2.2.5) [55, 56]. As depressive symptoms were rated on a 4-point Likert scale, the items were considered categorical indicators [57]. The mean- and variance-adjusted diagonal weighted least squares (WLSMV) estimator was

used in the analyses according to the software manual and recent literature [20, 58, 59]. As for missing values, pairwise deletion was used by default due to the use of the WLSMV estimator and the absence of external model covariates.

Multiple criteria were considered in order to evaluate the model fit. For the EFA and CFAs, the chi-square (χ^2) statistics, comparative fit index (CFI), Tucker-Lewis index (TLI) and root mean square error of approximation (RMSEA) were reported. The goodness of fit was assessed by the following combination of multiple criteria, with CFI and TLI > 0.95 and RMSEA < 0.06 for a relatively good fit [60]. As for the χ^2 statistics, they were presented here only for their use in calculating RMSEAs, not for evaluation of the model fit because they were sensitive to sample size [59].

To evaluate the measurement invariance, the changes of several indices were presented, consisting of changes in chi-square statistics ($\Delta\chi^2$), comparative fit index (ΔCFI), and root mean square error of approximation (ΔRMSEA). ΔCFI < 0.01 and ΔRMSEA < 0.015 indicated measurement invariance [61, 62].

Liu *et al. BMC Psychiatry*   (2023) 23:868

Page 6 of 12

Table 1 Model fit indices in the exploratory factor analysis

EFA Model	χ^2	df	χ^2/df	CFI	TLI	RMSEA (90% CI)
1-factor without correlated residual	2284.910[***]	20	114.246	0.772	0.681	0.191 (0.185 0.198)
1-factor with correlated residual	122.156[***]	19	6.429	0.990	0.985	0.042 (0.035 0.049)
2-factor without correlated residual	132.804[***]	13	10.216	0.988	0.974	0.055 (0.046 0.063)

[***] $p < 0.001$

Table 2 The estimated model parameters of one-factor model with correlated residual and two-factor model without correlated residual

Item number	Item content	1-factor with correlated residual	2-factor without correlated residual	
		DS	NS	DHF
6	I am in a low spirit	**0.75**	**0.76**	-0.02
7	I find it difficult to do anything	**0.63**	**0.65**	-0.04
11	I cannot sleep well	**0.53**	**0.53**	0.01
12	I feel happy	0.28	-0.01	**0.88**
14	I feel lonely	**0.78**	**0.76**	0.04
16	I have a happy life	0.32	0.09	**0.70**
18	I feel sad	**0.79**	**0.79**	0.00
20	I feel that I cannot continue with my life	**0.70**	**0.67**	0.07
Variance[a]		39.30%	52.66%	

[a] The content in the last row demonstrated explained percentage of the factor(s) in each model to the total variance

DS Depressive Symptoms, *NS* Negative Symptoms, *DHF* Diminished Happiness Feeling

Results

Exploratory factor analysis

Table 1 showed the model fit indices of three models, including one and two-factor model without correlated residual and one-factor model with correlated residual. The two-factor model with correlated residual was not suitable for the data because the residual covariance matrix was not positive definite, so its model fit indices were not provided. One-factor model without correlated residual yielded poor model fit. Both one-factor model with correlated residual and two-factor model without correlated residual had adequate model fit, with their estimated model parameters provided in Table 2.

In the one-factor model with correlated residual, the factor loadings of the two reverse-coded items were only 0.28 and 0.32, which were both below the cut-off value of 0.40 recommended by Worthington and Whittaker [63]. In addition, the model only explained 39.30% of the total variation in the sample. On the other hand, the two-factor model without correlated residual had no cross-loading items, with factor loadings ranging from 0.53 to 0.88, all above the cut-off value of 0.40. Furthermore, it explained 52.66% of the total variation in the sample. Hence, the two-factor model without correlated

residual (Model 5) was selected and used in the subsequent analyses.

In Model 5, the first factor, named Negative Symptoms, consisted of depressed affects (sad, low spirit, lonely, cannot continue) and somatic complains (difficult to do, sleep not well). With an eigenvalue of 2.940, it explained 36.75% of the total variation in the sample. The second factor, named Diminished Happiness Feeling, included two reverse-coded items relating to happiness feeling. With an eigenvalue of 1.273, it explained 15.91% of the total variation in the sample. The reliability coefficient for the Negative Symptoms factor was 0.74 (omega coefficient) and the coefficient for the Diminished Happiness Feeling factor was 0.71 (Spearman-Brown coefficient) [64].

Confirmatory factor analysis

The results of the CFAs were shown in Tables 3 and 4. Model 1 did not provide adequate fit at all. Model 2 considered both positive and negative wording effects, however, it could not be properly identified, which was consistent with previous literature [65]. Model 4 and Model 5 were equivalent models. The Diminished Happiness Feeling factor in Model 5 was replaced with the specification that two positive items loaded

Liu *et al. BMC Psychiatry*   (2023) 23:868

Page 7 of 12

Table 3 Fit indices of the competing models using the data at T2

Model	χ^2	df	χ^2/df	CFI	TLI	RMSEA (90% CI)
Model 1	1636.194[***]	19	86.115	0.826	0.743	0.207 (0.199 0.216)
Model 2[a]	—	—	—	—	—	—
Model 3	61.645[***]	14	4.403	0.995	0.990	0.041 (0.031 0.052)
Model 4	60.221[***]	19	3.170	0.996	0.993	0.033 (0.024 0.043)
Model 5	60.220[***]	19	3.169	0.996	0.993	0.033 (0.024 0.043)

[a] Model 2 could not be properly identified when applied to the data at T2

[***] $p < 0.001$

Table 4 Factor loadings of the competing models using the data at T2

Item	Model 1		Model 3		Model 4		Model 5	
	DM	SC	DS	Neg	DS	Pos	NS	DHF
n6	0.74		0.27	0.72	0.76		0.76	
n7		0.78	0.29	0.70	0.75		0.75	
n11		0.61	0.28	0.54	0.61		0.61	
n14	0.77		0.34	0.71	0.79		0.79	
n18	0.78		0.31	0.74	0.80		0.80	
n20		0.75	0.30	0.68	0.74		0.74	
p12	0.59		0.80		0.32	0.74		0.80
p16	0.61		0.82		0.33	0.74		0.82
AVE	0.49	0.51	0.23	0.47	0.44	0.55	0.55	0.66
Variance[a]	50.15%		58.30%		57.88%		57.95%	

[a] The content in the last row demonstrated explained percentage of the factor(s) in each model to the total variance

DM Depressed Mood, *SC* Somatic Complaints, *DS* Depressive Symptoms, *Neg* Negative Method, *Pos* Positive Method, *NS* Negative Symptoms, *DHF* Diminished Happiness Feeling, *AVE* Average variance extracted

simultaneously on the substantial factor and method factor in Model 4 [66]. Although Model 3, Model 4 and Model 5 demonstrated comparable fit and explained variance, the convergent validity of the substantive factor in Model 3 and Model 4, measured by Average variance extracted (AVE), was below the cut-off value 0.5 [67]. For Model 5, the AVEs of the two factors were both above the cut-off value, indicating good convergent validity. In the meanwhile, the correlation coefficient between the two factors in Model 5 was 0.40, whose square was much lower than the AVEs, demonstrating high discriminant validity. Therefore, from the statistical point of view, Model 5 could be selected as the most appropriate model.

From the substantive point of view, Model 5 suggested two substantive components, including Negative Symptoms factor and Diminished Happiness Feeling factor, which was preferred because (1) in the original article, Radloff argued that the positive items were used to break tendencies toward response set and evaluate positive affect [19]; (2) with depression, both the World Health Organization (WHO) and American Psychiatric Association (APA) considered that it involved depressed mood or loss of interest or pleasure, implying that diminished positive emotions was not just a wording effect but an important dimension [1, 68]; (3) in a broader perspective, the WHO noted that, "Mental health is an integral component of health and well-being and is more than the absence of mental disorder" [69]. In line with this, a dual-factor model of mental health including associated positive and negative factors was recommended to better explain mental health [70, 71].

Based on the above analysis, Model 5 was eventually selected and used in the subsequent measurement invariance test.

Measurement invariance

The model fit indices of the MG-CFAs were presented in Table 5. In the measurement invariance test across gender, all models in the hierarchy fitted well at both waves. The results showed that the model fit was not significantly deteriorated while imposing more and more strict constraints (including factor loadings, thresholds and residual variance) across groups, suggesting that the CES-D 8 measured the same construct for males and females at two time points.

Liu *et al. BMC Psychiatry*   (2023) 23:868

Page 8 of 12

Table 5 Model fit indices of the measurement invariance tests

model	χ^2	df	χ^2/df	$\Delta\chi^2$	RMSEA (90% CI)	CFI	TLI	ΔRMSEA	ΔCFI
Invariance test across gender at T1 (N = 3099, Male = 1648, Female = 1451)									
configural	145.374***	38	3.826		0.043 (0.035 0.050)	0.990	0.985		
metric	159.739***	44	3.630	16.371*	0.041 (0.034 0.048)	0.989	0.986	-0.001	-0.001
threshold	163.872***	58	2.825	19.229	0.034 (0.028 0.041)	0.990	0.990	-0.007	0.001
strict	238.184***	66	3.609	69.251***	0.041 (0.035 0.047)	0.983	0.986	0.007	-0.007
Invariance test across gender at T2 (N = 1978, Male = 1027, Female = 951)									
configural	76.472***	38	2.012		0.032 (0.021 0.042)	0.996	0.994		
metric	82.350***	44	1.872	6.989	0.030 (0.020 0.040)	0.996	0.995	-0.002	0.000
threshold	86.099**	58	1.484	11.853	0.022 (0.011 0.032)	0.997	0.997	-0.008	0.001
strict	124.283***	66	1.883	36.620***	0.030 (0.022 0.038)	0.994	0.995	0.008	-0.003
Invariance test over time (T1: N = 3099; T2: N = 1978)									
configural	191.674***	90	2.130		0.019 (0.015 0.023)	0.994	0.993		
metric	205.933***	96	2.145	15.016*	0.019 (0.016 0.023)	0.994	0.993	0.000	0.000
threshold	240.565***	110	2.187	37.118***	0.020 (0.016 0.023)	0.993	0.992	0.001	-0.001
strict	289.206***	118	2.451	49.229***	0.022 (0.018 0.025)	0.991	0.991	0.002	-0.002
Longitudinal cross-gender invariance test (T1: N = 3099, Male = 1648, Female = 1451; T2: N = 1978, Male = 1027, Female = 951)									
configural	322.486***	190	1.697		0.021 (0.017 0.025)	0.993	0.991		
metric	346.217***	208	1.665	27.102†	0.021 (0.017 0.024)	0.993	0.991	0.000	0.000
threshold	414.599***	250	1.658	75.547**	0.021 (0.017 0.024)	0.991	0.992	0.000	-0.002
strict	532.547***	266	2.002	122.085***	0.025 (0.022 0.029)	0.986	0.987	0.004	-0.005

† $p < 0.1$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$

In the longitudinal measurement invariance test, the changes in CFI and RMSEA indicated that strict invariance was supported over a two-year period. Moreover, the longitudinal cross-gender invariance was supported, demonstrating that the scale measured the same construct across gender over a two-year period. The structural invariance and actual gender differences and temporal differences in the CES-D 8 factor scores were also examined, and the results were provided in Tables S4 and S5 of the supplements.

Discussion

The current study aimed to provide more empirical evidence on the psychometric properties of the CES-D 8. A sample from the CFPS was used to identify the factor structure of the CES-D 8 among adolescents and examine its measurement invariance across gender and time (a two-year period). Previous literature focused primarily on general and aged adults, while few studies examined the factor structure and measurement invariance across gender and time among adolescents, especially in such a national sample. The study had three important findings.

First, based on the EFA and CFAs, a two-factor model was identified, including Negative Symptoms and Diminished Happiness Feeling. The Diminished Happiness Feeling factor contained two reverse-coded items describing happiness affect and the Negative Symptoms factor contained the other six items. The factor structure was similar to the previous results in Irish and South Africans [26, 72]. Although Adams et al. [26] made a slight modification to the two items, all the items ultimately loaded on "Negative Affect" factor and "Diminished Positive Affect" factor respectively. The Negative Symptoms factor in our model involved items from the "somatic complains" and "depressed affect" factors in Radloff's [19] original structure, and such integration had been observed in the studies conducted among Asians, Europeans and Africans, suggesting that depression might be characterized by some inherent mental and physical experiences across ages and cultures [25, 26]. It should also be noted that the two-factor model in current study was different from Karim's and Van de Velde's model in the two reverse-coded items, although they were equivalent models. That might be because their participants were aged adults, and anhedonia/loss of interest was more common in aged adults than adolescents [73, 74].

The results were inconsistent with those found in American samples [22, 24], where a different two-factor model was identified, including "depressed mood" and "somatic complaints" (i.e., Model 1). The most common explanation for this inconsistency was that Chinese people were more ashamed of reporting mental illness than westerners [75]. However, the same integration of

Liu *et al. BMC Psychiatry*   (2023) 23:868

Page 9 of 12

"depressed mood" and "somatic complaints" was also found among the Europeans [25]. There might exist other explanations, such as generation gap, as the American participants were significantly older than the European participants, or measurement bias, as dichotomous variables were used in the American studies while 4-categorical variables were used in the European studies. In summary, it is noteworthy that the construct of the depressive symptoms among the Chinese adolescents was not totally same with the Europeans and the Americans. More research was required to confirm the factor structure of the CES-D 8 among populations at different ages and varied cultures.

Second, based on our two-factor model, strict invariance across gender was supported, indicating that the construct (depressive symptoms) measured by the CES-D 8 was reliable, and the latent means and manifest means could be compared meaningly between girls and boys. This finding was consistent with the previous literature involving measurement invariance across gender, although their participants were young adults or aged adults, which might suggest that the CES-D 8 had comparable cross-gender stability across age groups [25, 27, 38].

Third, the longitudinal measurement invariance test suggested that strict invariance was supported in a two-year period among adolescents, even across gender. To our knowledge, although the CES-D 8 has been applied in longitudinal studies, this is the first research on longitudinal properties of the scale, especially across gender simultaneously [76]. Therefore, our findings of longitudinal strict invariance of the CES-D 8 extends its utility in terms of the longitudinal research.

Put it all together, the CES-D 8 is a suitable instrument for measuring depressive symptoms among adolescents. The brevity makes it preferable for large-scale administration to screen out the adolescents with relatively high depressive symptoms at a lower cost; in the meanwhile, it can guarantee a relatively robust data quality since it relieves the respondent burden due to adolescents' lack of attention [28, 43]. Furthermore, the measurement invariance test provided empirical evidence for the stability of the scale among adolescents, implicating the meaningful comparisons across gender or true changes in the development of depressive symptoms.

In spite of the strengths, there are four limitations in this study. First, the findings are based on an exclusive Chinese adolescent sample, so the generalizability of the CES-D 8 was not examined. Racial/ethnic generalizability is critical to any of the psychiatry measures [58]. In aged adults, different factor structures of the CES-D 8 had been found among Americans, Europeans and Africans [22, 24–27]. However, little research has been conducted to examine the psychometric properties of the scale in adolescents. More research should be conducted among diverse cohorts in different cultures in order to reach a more pervasive conclusion.

Second, despite the longitudinal design of the current study, data was only collected at two waves over a two-year period. In future research, more data at three waves or more over a longer period should be obtained. Using these data, not only the stability of the CES-D 8 can be examined more deeply, but also a latent growth model (LGM) can be established [77].The LGM can describe the developmental trajectories of the depressive symptoms over time, identify the intra-individual and inter-individual variability in reference levels and trajectories, and examine the different contributions of some protective and risk factors to the reference levels and trajectories.

Third, the CES-D 8 was self-reported and the only instrument used to measure depressive symptoms in the current study. It would be more reliable if interviews and/or other-reported instruments are combined. Furthermore, reliable interview instruments, such as WHO-Composite International Diagnostic Interview (CIDI), can be treated as a temporary "gold standard", allowing analysis of the performance of the CES-D 8 [78, 79]. The performance includes its sensitivity (ability to correctly identify patients), specificity (ability to correctly identify non-patients) and receiver operating characteristics (ROC) curves (used to establish an appropriate cut-off value to distinguish patients from non-patients) [80].

Fourth, although the possible common method bias caused by the wording effects had been considered in the EFA and CFA, the Diminished Happiness Feeling factor of the final model was the mix of substantial and method components. It would be more reliable if additional variables are introduced into the research and permit more advanced methods to identify and control method bias, such as confirmatory factor analysis marker technique and IV (i.e., independent variable) technique [81, 82].

Conclusions

This study reveals that the CES-D 8 remains reliable and stable across gender and over a two-year period among adolescents. The findings extend the related literature from general population or aged adults to adolescents, and from cross-sectional designs to longitudinal ones, indicating that it is a promising instrument to screen depressive symptoms among adolescents, especially in large-scale and longitudinal surveys.

Liu *et al. BMC Psychiatry*   (2023) 23:868

Page 10 of 12

Abbreviations

CES-D	Center for Epidemiologic Studies Depression Scale
CFPS	China Family Panel Studies
BDI	Beck Depression Inventory
PHQ-9	Patient Health Questionnaire-9
AHEAD	Asset and Health Dynamics Among the Oldest Old
HRS	Health and Retirement Study
ESS	European Social Survey
EFA	Exploratory factor analysis
CFA	Confirmatory factor analysis
ISSS	Institute of Social Science Survey
MG-CFA	Multi-group CFA
WLSMV	Mean- and variance-adjusted diagonal weighted least squares
CFI	Comparative fit index
TLI	Tucker-Lewis index
RMSEA	Root mean square error of approximation
IV	Independent variable
AVE	Average variance extracted

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12888-023-05316-4.

Additional file 1: Table S1. Parallel analysis of the data at T1. **Table S2.** Fit indices of the competing models using the data at T1. **Table S3.** Factor loadings of the competing models using the data at T1. **Table S4.** Model fit indices of the structural invariance tests. **Table S5.** actual gender and temporal differences in the CES-D 8 factor scores.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the China Family Panel Studies repository, http://www.isss.pku.edu.cn/cfps/. The codes were available at OSF HOME, https://osf.io/e9jqg/.

Declarations

Ethics approval and consent to participate

All methods performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 2013 Helsinki Declaration and its later amendments or comparable ethical standards.

The CFPS was approved by the Biomedical Ethics Review Committee of Peking University, and all participants were required to provide written informed consent. The ethical approval number was IRB00001052-14010.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. World Health Organization. Depression. https://www.who.int/news-room/fact-sheets/detail/depression. Accessed 2 May 2023.
2. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet. 2021;398(10312):1700–12. https://doi.org/10.1016/s0140-6736(21)02143-7.
3. Smith MV, Mazure CM. Mental health and wealth: depression, gender, poverty, and parenting. Annu Rev Clin Psychol. 2021;17:181–205. https://doi.org/10.1146/annurev-clinpsy-071219-022710.
4. Moreno-Agostino D, Wu YT, Daskalopoulou C, Hasan MT, Huisman M, Prina M. Global trends in the prevalence and incidence of depression: a systematic review and meta-analysis. J Affect Disord. 2021;281:235–43. https://doi.org/10.1016/j.jad.2020.12.035.
5. Weinberger AH, Gbedemah M, Martinez AM, Nash D, Galea S, Goodwin RD. Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. Psychol Med. 2018;48(8):1308–15. https://doi.org/10.1017/S0033291717002781.
6. Miller L, Campo JV. Depression in adolescents. N Engl J Med. 2021;385(5):445–9. https://doi.org/10.1056/NEJMra2033475.
7. Thapar A, Eyre O, Patel V, Brent D. Depression in young people. Lancet. 2022;400(10352):617–31. https://doi.org/10.1016/S0140-6736(22)01012-1.
8. Blakemore SJ. Adolescence and mental health. Lancet. 2019;393(10185):2030–1. https://doi.org/10.1016/S0140-6736(19)31013-X.
9. Costello EJ, Maughan B. Annual research review: optimal outcomes of child and adolescent mental illness. J Child Psychol Psychiatry. 2015;56(3):324–41. https://doi.org/10.1111/jcpp.12371.
10. Johnson D, Dupuis G, Piche J, Clayborne Z, Colman I. Adult mental health outcomes of adolescent depression: a systematic review. Depress Anxiety. 2018;35(8):700–16. https://doi.org/10.1002/da.22777.
11. Tariq A, Reid C, Chan SWY. A meta-analysis of the relationship between early maladaptive schemas and depression in adolescence and young adulthood. Psychol Med. 2021;51(8):1233–48. https://doi.org/10.1017/S0033291721001458.
12. Colman I, Wadsworth MEJ, Croudace TJ, Jones PB. Forty-year psychiatric outcomes following assessment for internalizing disorder in adolescence. Am J Psychiat. 2007;164(1):126–33. https://doi.org/10.1176/appi.ajp.164.1.126.
13. Copeland WE, Alaie I, Jonsson U, Shanahan L. Associations of childhood and adolescent depression with adult psychiatric and functional outcomes. J Am Acad Child Adolesc Psychiatry. 2021;60(5):604–11. https://doi.org/10.1016/j.jaac.2020.07.895.
14. Davey CG, McGorry PD. Early intervention for depression in young people: a blind spot in mental health care. Lancet Psychiatry. 2019;6(3):267–72. https://doi.org/10.1016/S2215-0366(18)30292-X.
15. Reangsing C, Punsuwun S, Schneider JK. Effects of mindfulness interventions on depressive symptoms in adolescents: a meta-analysis. Int J Nurs Stud. 2021;115: 103848. https://doi.org/10.1016/j.ijnurstu.2020.103848.
16. McDowell I. Measuring health: a guide to rating scales and questionnaires. 3rd ed. New York: Oxford University Press; 2006.
17. Kroenke K, Spitzer RL, Williams JBW. 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.
18. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry. 1961;4(6):561–71. https://doi.org/10.1001/archpsyc.1961.01710120031004.
19. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas. 1977;1(3):385–401. https://doi.org/10.1177/014662167700100306.
20. Niu L, He J, Cheng C, Yi J, Wang X, Yao S. Factor structure and measurement invariance of the Chinese version of the Center for Epidemiological Studies Depression (CES-D) scale among undergraduates and clinical patients. BMC Psychiatry. 2021;21:463. https://doi.org/10.1186/s12888-021-03474-x.

Liu *et al. BMC Psychiatry*   (2023) 23:868

Page 11 of 12

21. Yang W, Xiong G, Garrido LE, Zhang JX, Wang M-C, Wang C. Factor structure and criterion validity across the full scale and ten short forms of the CES-D among Chinese adolescents. Psychol Assess. 2018;30(9):1186–98. https://doi.org/10.1037/pas0000559.

22. Turvey CL, Wallace RB, Herzog R. A revised CES-D measure of depressive symptoms and a DSM-based measure of major depressive episodes in the elderly. Int Psychogeriatr. 1999;11(2):139–48. https://doi.org/10.1017/s1041610299005694.

23. Soldo BJ, Hurd MD, Rodgers WL, Wallace RB. Asset and health dynamics among the oldest old: an overview of the AHEAD study. J Gerontol Ser B-Psychol Sci Soc Sci. 1997;52:1–20. https://doi.org/10.1093/geronb/52B.Special_Issue.1.

24. Steffick, DE. Evaluation of the measures and data quality. In: Documentation of affective functioning measures in the Health and Retirement Study. Institute for Social Research, University of Michigan. 2000. https://hrs.isr.umich.edu/sites/default/files/biblio/dr-005.pdf. Accessed 12 May 2022.

25. Karim J, Weisz R, Bibi Z, Rehman SU. Validation of the eight-item Center for Epidemiologic Studies Depression Scale (CES-D) among older adults. Curr Psychol. 2015;34(4):681–92. https://doi.org/10.1007/s12144-014-9281-y.

26. Adams LB, Farrell M, Mall S, Mahlalela N, Berkman L. Dimensionality and differential item endorsement of depressive symptoms among aging Black populations in South Africa: findings from the HAALSI study. J Affect Disord. 2020;277:850–6. https://doi.org/10.1016/j.jad.2020.08.073.

27. Van de Velde S, Levecque K, Bracke P. Measurement equivalence of the CES-D 8 in the general population in Belgium: a gender perspective. Arch PUblic Health. 2009;67:15. https://doi.org/10.1186/0778-7367-67-1-15.

28. Hoyer RS, Elshafei H, Hemmerlin J, Bouet R, Bidet-Caulet A. Why are children so distractible? Development of attention and motor control from childhood to adulthood. Child Dev. 2021;92(4):e716–37. https://doi.org/10.1111/cdev.13561.

29. Hyde JS, Mezulis AH, Abramson LY. The ABCs of depression: integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. Psychol Rev. 2008;115(2):291–313. https://doi.org/10.1037/0033-295X.115.2.291.

30. Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. J Abnorm Psychol. 1998;107(1):128–40. https://doi.org/10.1037/0021-843x.107.1.128.

31. Platt JM, Bates L, Jager J, McLaughlin KA, Keyes KM. Is the US gender gap in depression changing over time? A meta-regression Am J Epidemiol. 2021;190(7):1190–206. https://doi.org/10.1093/aje/kwab002.

32. Girgus JS, Yang K. Gender and depression. Curr Opin Psychol. 2015;4:53–60. https://doi.org/10.1016/j.copsyc.2015.01.019.

33. Galambos NL, Barker ET, Krahn HJ. Depression, self-esteem, and anger in emerging adulthood: seven-year trajectories. Dev Psychol. 2006;42(2):350–65. https://doi.org/10.1037/0012-1649.42.2.350.

34. Schubert KO, Clark SR, Van LK, Collinson JL, Baune BT. Depressive symptom trajectories in late adolescence and early adulthood: a systematic review. Aust N Z J Psych. 2017;51(5):477–99. https://doi.org/10.1177/0004867417700274.

35. 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.

36. Van de Velde S, Bracke P, Levecque K, Meuleman B. Gender differences in depression in 25 European countries after eliminating measurement bias in the CES-D 8. Soc Sci Res. 2010;39(3):396–404. https://doi.org/10.1016/j.ssresearch.2010.01.002.

37. Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJ, et al. Exploratory structural equation modeling, Integrating CFA and EFA: application to Students' Evaluations of University Teaching. Struct Equ Modeling. 2009;16(3):439–76. https://doi.org/10.1080/10705510903008220.

38. Klusáček J, Kudrnáčová M, Soukup P. Validation of CES-D8 among Czech university students during COVID-19 pandemic. Cesk Psychol. 2022;66(4):398–415. https://doi.org/10.51561/cspsych.66.4.398.

39. Costello DM, Swendsen J, Rose JS, Dierker LC. Risk and protective factors associated with trajectories of depressed mood from adolescence to early adulthood. J Consult Clin Psychol. 2008;76(2):173–83. https://doi.org/10.1037/0022-006X.76.2.173.

40. Kwong ASF, López-López JA, Hammerton G, Manley D, Timpson NJ, Leckie G, et al. Genetic and environmental risk factors associated with trajectories of depression symptoms from adolescence to young adulthood. JAMA Netw Open. 2019;2(6):e196587. https://doi.org/10.1001/jamanetworkopen.2019.6587.

41. Widaman KF, Ferrer E, Conger RD. Factorial invariance within longitudinal structural equation models: measuring the same construct across time. Child Develop Perspect. 2010;4(1):10–8. https://doi.org/10.1111/j.1750-8606.2009.00110.x.

42. Grouzet FM, Otis N, Pelletier LG. Longitudinal cross-gender factorial invariance of the Academic Motivation Scale. Struct Equ Modeling. 2006;13(1):73–98. https://doi.org/10.1207/s15328007sem1301_4.

43. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cogn Psychol. 1991;5(3):213–36. https://doi.org/10.1002/acp.2350050305.

44. Xie Y, Hu J. An introduction to the China Family Panel Studies (CFPS). Chin Sociol Rev. 2014;47(1):3–29. https://doi.org/10.2753/csa2162-0555470101.

45. Wu Q, Dai L, Zhen Q, Gu L, Wang Y. The variables in cross-year library. In: CFPS-40: cleaning report of the cross-year library of the individual core variables. Institute of Social Science Survey, Peking University. 2021. http://www.isss.pku.edu.cn/cfps/docs/20210416110205277261.pdf. Accessed 15 Oct 2022.

46. Asparouhov T, Muthén B. Exploratory structural equation modeling. Struct Equ Modeling. 2009;16(3):397–438. https://doi.org/10.1080/10705510903008204.

47. Muthén LK, Muthén BO. Mplus user's guide, seventh ed. 1998–2015. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf. Accessed 10 Jun 2022.

48. Marsh HW, Scalas LF, Nagengast B. Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: traits, ephemeral artifacts, and stable response styles. Psychol Assess. 2010;22(2):366–81. https://doi.org/10.1037/a0019225.

49. DiStefano C, Motl RW. Further investigating method effects associated with negatively worded items on self-report surveys. Struct Equ Modeling. 2006;13(3):440–64. https://doi.org/10.1207/s15328007sem1303_6.

50. Reise SP. The rediscovery of bifactor measurement models. Multivariate Behav Res. 2012;47(5):667–96. https://doi.org/10.1080/00273171.2012.715555.

51. Ou XC. Multidimensional structure or wording effect? Reexamination of the factor structure of the Chinese General Self-Efficacy Scale. J Pers Assess. 2022;104(1):64–73. https://doi.org/10.1080/00223891.2021.1912059.

52. Millsap RE, Yun-Tein J. Assessing factorial invariance in ordered-categorical measures. Multivariate Behav Res. 2004;39(3):479–515. https://doi.org/10.1207/s15327906mbr3903_4.

53. Liu Y, Millsap RE, West SG, Tein J-Y, Tanaka R, Grimm KJ. Testing measurement invariance in longitudinal data with ordered-categorical measures. Psychol Methods. 2017;22(3):486–506. https://doi.org/10.1037/met0000075.

54. Guo B, Kaylor-Hughes C, Garland A, Nixon N, Sweeney T, Simpson S, et al. Factor structure and longitudinal measurement invariance of PHQ-9 for specialist mental health care patients with persistent major depressive disorder: Exploratory Structural Equation Modelling. J Affect Disord. 2017;219:1–8. https://doi.org/10.1016/j.jad.2017.05.020.

55. R Core Team. R: a language and environment for statistical computing (version 4.2.1). 2022; https://www.R-project.org/.

56. Revelle W. Psych: procedures for psychological, psychometric, and personality research (version 2.2.5). 2022; https://CRAN.R-project.org/package=psych.

57. Johnson DR, Creech JC. Ordinal measures in multiple indicator models: a simulation study of categorization error. Am Sociol Rev. 1983;48(3):398–407. https://doi.org/10.2307/2095231.

58. Dong L, Wu H, Waldman ID. Measurement and structural invariance of the Antisocial Process Screening Device. Psychol Assess. 2014;26(2):598–608. https://doi.org/10.1037/a0035139.

59. Pendergast LL, von der Embse N, Kilgus SP, Eklund KR. Measurement equivalence: a non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. J Sch Psychol. 2017;60:65–82. https://doi.org/10.1016/j.jsp.2016.11.002.

60. Hu L-t, 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.
61. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model. 2002;9(2):233–55. https://doi.org/10.1207/s15328007sem0902_5.
62. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Model. 2007;14(3):464–504. https://doi.org/10.1080/10705510701301834.
63. Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. Couns Psychol. 2006;34(6):806–38. https://doi.org/10.1177/0011000006288127.
64. Eisinga R, Grotenhuis MT, Pelzer B. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? Int J Public Health. 2013;58:637–42. https://doi.org/10.1007/s00038-012-0416-3.
65. Kenny DA, Kashy DA. Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. Psychol Bull. 1992;112(1):165–72.
66. Kline RB. Principles and practice of structural equation modeling. 4th ed. New York: Guilford publications; 2015.
67. Hair J, Black W, Babin B, Anderson R. Multivariate data analysis. 7th ed. Upper Saddle River: Prentice-Hall; 2010.
68. American Psychiatric Association, DSM-5 Task Force. Diagnostic and statistical manual of mental disorders. 5th ed (DSM-5). Washington, DC: American Psychiatric Association Publishing; 2013.
69. World Health Organization. World mental health report: transforming mental health for all. 2022. https://apps.who.int/iris/rest/bitstreams/1433523/retrieve. Accessed 10 Apr 2023.
70. Suldo SM, Shaffer EJ. Looking beyond psychopathology: the dual-factor model of mental health in youth. Sch Psychol Rev. 2008;37(1):52–68. https://doi.org/10.1080/02796015.2008.12087908.
71. Clark KN, Malecki CK. Adolescent mental health profiles through a latent dual-factor approach. J Sch Psychol. 2022;91:112–28. https://doi.org/10.1016/j.jsp.2022.01.003.
72. Briggs R, Carey D, O'Halloran AM, Kenny RA, Kennelly SP. Validation of the 8-item Centre for Epidemiological Studies Depression Scale in a cohort of community-dwelling older people: data from The Irish Longitudinal Study on Ageing (TILDA). Eur Geriatr Med. 2018;9(1):121–6. https://doi.org/10.1007/s41999-017-0016-0.
73. Fiske A, Wetherell JL, Gatz M. Depression in older adults. Annu Rev Clin Psychol. 2009;5:363–89. https://doi.org/10.1146/annurev.clinpsy.032408.153621.
74. Rice F, Riglin L, Lomax T, Souter E, Potter R, Smith DJ, et al. Adolescent and adult differences in major depression symptom profiles. J Affect Disord. 2019;243:175–81. https://doi.org/10.1016/j.jad.2018.09.015.
75. Parker G, Gladstone G, Chee KT. Depression in the planet's largest ethnic group: the Chinese. Am J Psychiat. 2001;158(6):857–64. https://doi.org/10.1176/appi.ajp.158.6.857.
76. Turvey CL, Schultz SK, Beglinger L, Klein DM. A longitudinal community-based study of chronic illness, cognitive and physical function, and depression. Am J Geriatr Psychiatr. 2009;17(8):632–41. https://doi.org/10.1097/JGP.0b013e31819c498c.
77. Hancock GR, Harring JR, Lawrence FR. Using latent growth modeling to evaluate longitudinal change. In: Hancock GR, Mueller RO, editors. Structural equation modeling: a second course. 2nd ed. Charlotte: Information Age Publishing; 2013. p. 309–42.
78. Dang L, Dong L, Mezuk B. Shades of blue and gray: a comparison of the Center for Epidemiologic Studies Depression Scale and the Composite International Diagnostic Interview for assessment of depression syndrome in later life. Gerontologist. 2020;60(4):e242–53. https://doi.org/10.1093/geront/gnz044.
79. Wittchen HU. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. J Psychiat Res. 1994;28(1):57–84. https://doi.org/10.1016/0022-3956(94)90036-1.
80. Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatr. 2007;96(5):644–7. https://doi.org/10.1111/j.1651-2227.2006.00178.x.
81. Williams LJ, Hartman N, Cavazotte F. Method variance and marker variables: a review and comprehensive CFA marker technique. Organ Res Methods. 2010;13(3):477–514. https://doi.org/10.1177/1094428110366036.
82. Jordan PJ, Troth AC. Common method bias in applied settings: the dilemma of researching in organizations. Aust J Manag. 2020;45(1):3–14. https://doi.org/10.1177/0312896219871976.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.