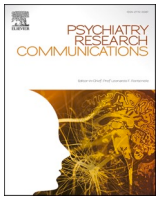
Psychiatry Research Communications 4 (2024) 100196

Contents lists available at ScienceDirect

Psychiatry Research Communications

journal homepage: www.sciencedirect.com/journal/Psychiatry-Research-Communications



Factor structure of the CES-D-20 scale in young adults in Singapore: An examination of six alternative structures and measurement invariance by sex

Nadyanna M. Majeed [a,*], K.T.A. Sandeeshwara Kasturiratna [b], Nicole R.Y. Chen [b], Andree Hartanto [b,**]

[a] Faculty of Arts and Social Sciences, National University of Singapore, Singapore
[b] School of Social Sciences, Singapore Management University, Singapore

ARTICLE INFO

Handling Editor: Dr. Leonardo Fontenelle

Keywords:
CES-D-20
Depression
Factor structure
Measurement invariance
Latent variable analysis

ABSTRACT

Building on past research that has indicated cultural variations in the factor structure of the 20-item Center for Epidemiologic Studies Depression Scale (CES-D-20), as well as several limitations validating the CES-D-20 within a Singapore context, this study investigated the CES-D-20's factor structure and sex-based measurement invariance in Singapore. A large young adult sample in Singapore ($N = 1008$) was utilised, and the original CES-D-20 was administered to all participants. The findings confirmed the four-factor model's robustness over five other alternative models (RMSEA = .05, SRMR = .04), where each of the 20 items loaded significantly onto its respective factor (all loadings≥.42, all $ps < .001$). All four factors were also significantly inter-correlated to a medium to very large extent (all $ps < .001$). Configural and weak invariance were established across sexes; however, strong invariance required modification, where the release of a single intercept constraint resulted in strong partial invariance, implying sex differences in the expression of crying. Homogeneity tests suggested no significant sex differences in latent factor means, variances, and covariances. The importance of cultural- and sex-specific considerations when utilising the CES-D-20 are discussed.

1. Introduction

The Center for Epidemiologic Studies Depression Scale (CES-D-20) is a critical tool widely utilised in psychological research and clinical practice to assess depressive symptoms (Blodgett et al., 2021; Mitchell and Coyne, 2007; Santor et al., 2006; Vilagut et al., 2016). Developed initially by Radloff (1977), the CES-D-20 is a self-report scale that helps identify individuals at risk for depression, meant to be applied across diverse age groups, countries, and in both community and institutionalised samples (Blodgett et al., 2021). Due to its relative brevity and accessibility in the public domain, the CES-D-20 has established itself as a foundational instrument in mental health research and clinical practice (Mitchell and Coyne, 2007; Ramírez-Vélez et al., 2023; Vilagut et al., 2016).

While the CES-D-20 was initially structured to encapsulate four distinct factors—depressive affect, positive affect, somatic symptoms, and interpersonal symptoms (Radloff, 1977)—the universality of this factorial arrangement has been frequently scrutinised (e.g., Carleton et al., 2013; Thorson and Powell, 1993). Multiple alternate factor solutions have been presented, including two-factor solutions (Rivera-Medina et al., 2010; Schroevers et al., 2000), three-factor solutions (Carleton et al., 2013; Wang et al., 2013), as well as a five-factor solution (Thorson and Powell, 1993). To reconcile these differences, Cosco et al. (2017) conducted a comprehensive evaluation of the CES-D-20 by comparing five different factor models (one-factor model, two-factor model, two three-factor models, and the original four-factor model) in a relatively large sample. Their research reaffirmed the robustness of the original four-factor structure proposed by Radloff (1977). However, literature suggests that the factor structure of the CES-D-20 may vary significantly across different cultural backgrounds, emphasising the need for localised validations (Blodgett et al., 2021; Kim et al., 2011). Indeed, a meta-analysis investigating the factor structure of CES-D-20 found that US-based studies were more likely to be consistent with the original four-factor structure, while studies based in

* Corresponding author. National University of Singapore, Faculty of Arts and Social Sciences, 5 Arts Link, Singapore 117570, Singapore.
** Corresponding author. Singapore Management University, School of Social Sciences, 10 Canning Rise, Level 5, Singapore, 179873.
E-mail addresses: nadyanna@u.nus.edu (N.M. Majeed), andreeh@smu.edu.sg (A. Hartanto).

https://doi.org/10.1016/j.psycom.2024.100196
Received 3 June 2024; Received in revised form 2 September 2024; Accepted 1 October 2024
Available online 10 October 2024

Asia predominantly reported two- or three-factor structures (Blodgett et al., 2021). This underlines the necessity of examining the scale's applicability in a multicultural landscape such as Singapore.

Previous efforts to validate the CES-D-20 in Singapore have faced limitations. These include restricted demographic groups (e.g., caregivers of those with dementia; Ying et al., 2019), and a lack of comprehensive evaluation of the scale's multidimensionality (e.g., Stahl et al., 2008). While some studies have made efforts to validate the shorter version of the CES-D (i.e., CES-D-10) in Singapore (e.g., Lee and Chokkanathan, 2008; Mohebbi et al., 2018), the 10-item and 20-item measures are not necessarily comparable. For example, the 10-item measure has been consistently found to only capture one or two factors, instead of the full four-factor structure (Björgvinsson et al., 2013; González et al., 2017; Mohebbi et al., 2018). Such gaps hinder the accurate assessment of depressive symptoms and impede the development of effective mental health interventions tailored to the local populace.

Furthermore, an additional research gap exists in the exploration of measurement invariance across sexes in the Singaporean context. Depression manifests differently between males and females, influenced by both biological and psychosocial factors, which may affect the psychometric properties of tools like the CES-D-20 (Barbosa-Leiker et al., 2021). These sex and/or gender differences can be attributed to variations in hormonal fluctuations, socialisation processes, and coping strategies employed by males and females when confronted with stress and depression (Carle et al., 2008; Girgus and Yang, 2015; Hyde and Mezulis, 2020). Studies on other instruments, such as the Beck Depression Inventory, have shown significant differences in factor structures across sexes, indicating potential biases if not appropriately addressed (Wu, 2010).

Given these contexts, our study aims to address these critical gaps by (1) evaluating the factor structure of the CES-D-20 in a large sample of young adults in Singapore by comparing the six different factor structures previously proposed in the literature, and by (2) examining the measurement invariance of the best-fitting factor structure from the previous step across male and female individuals. This approach not only seeks to validate the use of the CES-D-20 in Singapore but also enhances our understanding of how depressive symptoms are expressed and interrelated across different sexes in this unique cultural setting.

2. Method

2.1. Sample

A total of 1010 unique individuals in Singapore provided data across four separate periods of data collection from December 2020 to March 2023 as part of a larger set of studies examining various personality traits (Chen et al., 2024). Individuals were eligible to participate as long as they were studying locally at the time of data collection; no demographic or health restrictions were applied. A fixed sample size was not predetermined; recruitment was carried out over a fixed timeframe during each of the four periods and any eligible individual who expressed interest was allowed to participate. In order to ensure that all data collected was independent, individuals were not allowed to participate multiple times; each individual provided a response during only one of the four periods. All participants provided informed consent, and all participants were compensated for their time. Data collection was approved by the Institutional Review Board at the last author's university. Of the 1010 participants, two did not provide complete data, and were thus removed from the analytic sample in the current work. All descriptive and inferential statistics in the current work were thus drawn from the remaining sample of 1008 participants (82.44% Singaporean by nationality).[1] Sample characteristics are presented in Table 1.

2.2. Measure and procedure

The original CES-D-20 (Radloff, 1977) was administered to all participants. It attempts to measure the frequency of experiencing each of 20 depressive symptoms, with the timeframe specified as "during the past week" (e.g., "During the past week, I felt sad", "During the past week, I talked less than usual"). Each symptom is rated on a 4-point scale (0 = *Rarely or none of the time (Less than 1 day)*, 1 = *Some or a little of the time (1–2 days)*, 2 = *Occasionally or a moderate amount of time (3–4 days)*, 3 = *Most or all of the time (5–7 days)*).

Each participant responded to the CES-D-20 within a single sitting, on their own personal laptops or desktops. Each data collection session was run either in-person in a closed room with multiple participants, or virtually over Zoom with multiple participants. Participants were not able to see each other's computer screens in order to maintain participant privacy. Every session was proctored by a trained research assistant in an effort to maximise data quality by preventing participants from being distracted. The research assistant(s) did not have access to participants' screens. All participants also provided their demographic data either in the same session or in a separate session one day earlier or one day later. All data was recorded using the survey platform Qualtrics.

2.3. Analytic plan

When determining the factors that a measure may tap onto, two possible approaches are usually chosen from: exploratory factor analysis and confirmatory factor analysis. Exploratory factor analysis is more appropriate when in the early stages of developing a new measure (e.g., when determining which items to retain) and when the measurement model is not known a priori (Hurley et al., 1997; Kelloway, 1995). However, in the context of the current work, there is a fixed set of six possible measurement models to be applied to the CES-D-20, each with its own theoretical basis. In addition, the CES-D-20 itself is an established and widely-used measure (Blodgett et al., 2021; Mitchell and Coyne, 2007; Santor et al., 2006; Vilagut et al., 2016). We emphasise

Table 1
Sample characteristics.

Characteristic	*M* (*SD*) or %	Observed Range
Age (years)	21.93 (1.82)	18–30
Sex (% female)	72.72%	
Nationality (% Singaporean)	82.44%	
Race (% Chinese, ethnic majority)	78.47%	
Objective socioeconomic status (monthly household income)		
Less than $2000	10.91%	
$2000–$5999	29.96%	
$6000–$9999	19.44%	
$10,000–$14,999	21.73%	
$15,000–$19,999	8.04%	
More than or equal to $20,000	9.92%	
Subjective socioeconomic status (self-rated ladder ranking)	6.16 (1.32)	2–10

Note. $N = 1008$. Subjective socioeconomic status was measured using a 10-point ladder scale adapted from Adler et al. (2000), with the bottom rung (1) representing the lowest status in the community and the top rung (10) representing the highest status in the community.

[1] All participants were assumed to be native or otherwise fluent speakers of English, as the language of education and language of administration in Singapore is English.

that, in the current work, we do not aim to modify the CES-D-20 measure (e.g., through rewording or dropping of items); we simply examine the possible factor solutions that can be drawn from the full existing 20-item scale, so that researchers can appropriately model data collected using the CES-D-20. As such, confirmatory factor analysis is the more relevant approach (Hurley et al., 1997; Williams, 1995).

Thus, we conducted confirmatory factor analysis to assess what factor structure provided the best fit to the data. In addition to the five different factor structures tested by Cosco et al. (2017) in their examination of the CES-D-20 administered to US adults, we also tested an additional factor structure proposed by Thorson and Powell (1993). We then extended upon previous works by testing for measurement invariance across the two sexes (female vs. male), due to potential sex differences in the manifestation of depression (Carle et al., 2008; Girgus and Yang, 2015; Hyde and Mezulis, 2020; Wu, 2010). Lastly, we computed estimates of internal consistency for each factor of the CES-D-20 that was identified in the previous steps.

2.3.1. Factor structure

We examined six different factor structures which have been proposed in the existing literature. We first modelled Sheehan et al.'s (1995) one-factor solution where all items loaded onto a single *depression* factor (Model 1). Then, we modelled Gomez and McLaren's (2015) two-factor solution where four items ("I felt I was just as good as other people", "I felt hopeful about the future", "I was happy", "I enjoyed life") loaded onto a *positive affect* factor and the remaining 16 items loaded onto a *general depression* factor (Model 2). Following that, we modelled two separate three-factor solutions: Guarnaccia et al.'s (1989) model which involves the same 4-item *positive affect* factor, a 2-item *interpersonal* factor ("People were unfriendly", "I felt that people disliked me"), and a remaining 14-item *somatic and depressive* factor (Model 3a); and Fifield and Reisine (1992) model which involves the same 2-item *interpersonal* factor as previously specified, a 7-item *somatic* factor ("I was bothered by things […]", "I did not feel like eating […]", "I had trouble keeping my mind on what I was doing", "[…] everything I did was an effort", "My sleep was restless", "I talked less than usual", "I could not get going"), and a remaining 11-item *affect* factor (Model 3b). Then, we modelled Radloff (1977) original four-factor solution: *depressive affect*, *positive affect*, *interpersonal*, *somatic* (Model 4). Lastly, we modelled Thorson and Powell's (1993) five-factor solution, which was based on the original four factors with an additional fifth factor of *self worth* made of three items ("I felt I was just as good as other people" from positive affect, "[…] my life had been a failure" from depressive affect, "I felt fearful" from depressive affect; Model 5). Where multiple factors were involved (i.e., all models except Model 1), all factors were allowed to freely inter-correlate.

Each model was evaluated independently in terms of its fit; models were considered to have good fit if RMSEA$\leq$.06, SRMR$\leq$.08, CFI$>$.95, and TLI$>$.95, following the analytic plan of Cosco et al. (2017).[2] Extending beyond Cosco et al. (2017) and in line with literature on best practices in model comparison and model selection (Lin et al., 2017; Lu et al., 2017; Chakrabarti and Ghosh, 2011; Huang, 2017; Vrieze, 2012), the six models were then compared with reference to their AIC and BIC values in order to select the most appropriate model. The model with the lowest AIC and lowest BIC was considered to be the most appropriate model.

2.3.2. Exploration of sex differences

2.3.2.1. Measurement invariance. Invariance of the measure across the two sexes (female and male) was tested for factor structure that displayed the best fit. Specifically, invariance was tested in four steps: configural invariance (whether the general factor structure was similar), weak invariance (whether the factor loadings were equal across sexes), strong invariance (whether the intercepts were equal across sexes), and strict invariance (whether the residuals and residual covariances were equal across sexes).

Model comparisons using AIC, BIC, and ΔCFI were conducted at each step in order to assess invariance.[3] Following conventional rules, in order for invariance to hold, AIC and BIC values should be lower (Cao and Liang, 2022) and the change in CFI should be less than .010 (i.e., ΔCFI$<$.010; Cheung, 2002). We considered invariance to be held if at least two of the three criterion were met. At each step, if full invariance did not hold, up to five of the newly-added constraints were released in order to improve model fit. Constraints were selected for release based on the score test (also known as the Lagrange Multiplier test; Bentler and Chou, 1992) for releasing constrained parameters in a model; the constraint with the lowest *p*-value was always selected. One constraint was released at a time, up to a maximum of five, until partial invariance held. As invariance testing is a sequential process, if at any step we did not achieve at least partial invariance, we discontinued testing and instead interpreted the final model that was available.

2.3.2.2. Homogeneity of latent depression by sex. In a final exploratory step, we tested in two separate steps for homogeneity across the sexes. Specifically, in the first step, we tested for equality of latent means (i.e., do males and females exhibit similar latent levels of each of the factors of depression), and in the second step, we tested for homogeneity of latent variances and covariances (i.e., do males and females exhibit similar variances and covariances of the factors of depression). In testing for homogeneity, we used the same criteria as that for measurement invariance testing. Violation of the criteria would imply that the two sexes were different on their latent means and/or latent variances and covariances.

2.3.3. Psychometric properties

Based on the final factor solution selected from the previous step, we calculated various estimates related to the psychometric properties of the CES-D-20. First, we calculated two estimates of (unidimensional) internal consistency (Cho, 2016) for each factor of the CES-D-20. Specifically, we estimated tau-equivalent reliability (i.e., Cronbach's α) and congeneric reliability (i.e., ω). Internal consistency was considered good if the reliability estimate was above .70 (Tavakol and Dennick, 2011), although some argue that reliability estimates of .60 may be sufficient (e.g., Rudner and Schafer, 2001). Second, we calculated the average variance extracted (i.e., AVE) for each factor. AVE should be at least .50 for a factor to be considered acceptable in sufficiently explaining its indicators (Cheung et al., 2024; Fornell and Larcker, 1981). In the case where AVE values fall below .50, the factor can still be considered to have adequate psychometric properties if ω exceeds .60 (Fornell and Larcker, 1981).

2.4. Transparency and openness

This study's design and its analysis plan were not pre-registered. Relevant data and analytic code will be made publicly available on Researchbox #2912 (https://researchbox.org/2912) upon publication. All analyses were conducted in R version 4.4.1 (R Core Team, 2023), using *lavaan* version .6–18 (Rosseel, 2012) and *semTools* version .5–6

[2] It should be noted that none of the five models achieved good fit on the CFI (maximum achieved = .94) and TLI (maximum achieved = .93) in Cosco et al.'s (2017) analysis, which used a sample of adults in the US.

[3] For completeness of reporting, we also reported the conventional χ^2 difference test. In order for invariance to hold, the χ^2 difference test should be statistically non-significant (Cao and Liang, 2022). However, it is noteworthy that the χ^2 difference test can yield statistically significant results even when the misfit is negligible, due to its sensitivity to large sample sizes (Meade et al., 2008).

(Jorgensen et al., 2022). Visualisations were created using *ggplot2* version 3.5.1 (Wickham, 2016), *lavaanPlot* version .8.1 (Lishinski, 2024), and *qgraph* version 1.9.8 (Epskamp et al., 2012).

3. Results

3.1. Factor structure

The one-factor solution displayed markedly poor fit (Table 2). Four out of six of the possible models displayed close to good fit (Model 2, Model 3a, Model 4, and Model 5; Table 2). A comparison of the six models' AIC and BIC values resulted in Model 4, the original four-factor solution, being selected as the best model (Table 2).

In Model 4, each of the 20 items loaded significantly onto its respective factor (all loadings≥.42, all *ps* < .001). In addition, all four factors were significantly inter-correlated to a medium to very large extent (all *ps* < .001; Table 3); in terms of absolute magnitude, the weakest correlation was found between the interpersonal and positive affect factors ($r = -.29$), while the strongest correlation was found between the depressive affect and somatic factors ($r = .89$).

3.2. Measurement invariance by sex

First, configural invariance was examined across the two sexes. The configural invariance model displayed close to good fit (RMSEA = .05, SRMR = .04, CFI = .94, TLI = .93) and was thus used for the next step of invariance testing.

Weak invariance was tested by comparing the previous configural invariance model with a new model with factor loadings constrained to be equal across the two sexes. Examinations of ΔCFI, AIC, and BIC values revealed that weak invariance held based on ΔCFI and BIC when all factor loadings were constrained (Table 4).

Similarly, strong invariance was tested by comparing the previous weak invariance model with a new model where intercepts were constrained to be equal across the two sexes. A further examination of ΔCFI, AIC, and BIC values revealed that strong invariance did not hold when all intercept constraints were included (Table 4). The release of a single intercept constraint (i.e., the "CRY" item) resulted in strong partial invariance according to ΔCFI and BIC.

We then tested for strict (partial) invariance by comparing the previous strong partial invariance model with a new model where residuals and residual covariances were constrained to be equal across the two sexes. We found that invariance based on the new model (i.e., strict invariance with only one intercept released) was held based on ΔCFI and BIC (Table 4).

3.3. Homogeneity of latent means, variances, and covariances by sex

Using the strict invariance model with the "CRY" intercept released (i.e., allowed to differ across the sexes) with the additional constraint of equal latent means, the first test of homogeneity showed no evidence that males and females differed significantly in terms of their latent mean levels of the four depression factors, based on ΔCFI, AIC, and BIC consistently. The second test of homogeneity (i.e., with the additional constraint of equal latent variances and covariances) also showed no evidence that males and females differed significantly in terms of their latent variances and covariances of the four depression factors, based on ΔCFI and BIC.

Table 2
Model fit indices and model comparison indices of the various factor structures.

Model	RMSEA	SRMR	CFI	TLI	AIC	BIC
Threshold	≤.06	≤.08	>.95	>.95	Lowest	Lowest
Model 1 (one-factor)	.10	.09	.76	.73	45636	45833
Model 2 (two-factor)	**.06**	**.04**	.93	.92	44358	44559
Model 3a (three-factor)	**.05**	**.04**	.94	.93	44287	44498
Model 3b (three-factor)	.10	**.08**	.78	.75	45496	45707
Model 4 (four-factor)	**.05**	**.04**	.95	.94	**44227**	**44453**
Model 5 (five-factor)	**.06**	**.06**	.92	.91	44414	44660

Note. $N = 1008$. Bolded values indicate good fit according to the corresponding index.

Table 3
Inter-correlations between the four factors of CES-D-20.

	1	2	3	4
1. Depressive affect	–			
2. Interpersonal	.76	–		
3. Positive affect	−.34	−.29	–	
4. Somatic	.89	.73	−.31	–

Note. $N = 1008$. All inter-correlations were statistically significant (all *ps* < .001).

Table 4
Model comparisons for measurement invariance by sex.

Model	$\Delta\chi^2$	Δdf	*p*	ΔCFI	AIC	BIC
Threshold				<.010	Lowest	Lowest
Measurement invariance						
Configural					44122	44771
Weak (vs. configural)	85.30	16	<.001	.009	44176	44746
Strong (vs. weak)						
Full	107.00	16	<.001	.012	44250	44742
Partial (one intercept released)	68.80	15	<.001	.007	44214	44711
Strict (vs. strong partial)	57.50	20	<.001	.005	44232	44630
Homogeneity						
Means	8.23	4	.084	.001	44232	44611
Variances and covariances	21.80	10	.016	.002	44234	44563

Note. $N = 1008$ (female $N = 733$, male $N = 275$). Homogeneity tests were conducted based on the strict invariance model with one intercept ("CRY" item) released.

3.4. Factor structure by sex

As strict invariance as well as full homogeneity held across the two sexes except for the "CRY" intercept, all results were in line with what was previously found in the overall model. Fig. 1 depicts the factor loadings and latent inter-correlations (equal across both sexes). In terms of the "CRY" intercept which was allowed to vary between the two sexes, we found that it was significantly higher in females (standardised estimate = .72) than in males (standardised estimate = .38). This implies that, all else being equal (e.g., a male participant and a female participant of comparable levels on all latent depression factors), a female participant would have higher levels of "CRY" than would a male participant.

3.5. Psychometric properties

The psychometric properties of the four factors of the CES-D-20 in the current sample were generally good (Table 5). Across the four factors, across both males and females, both α and ω estimates generally exceeded the .70 threshold. Of note, all reliability estimates exceeded the more relaxed .60 threshold. Separately, three of the four factors exhibited acceptable levels of AVE, while the somatic factor did not (Table 5). Despite the low AVE of the somatic factor, the somatic factor was still considered psychometrically adequate given that its ω exceeded .60.

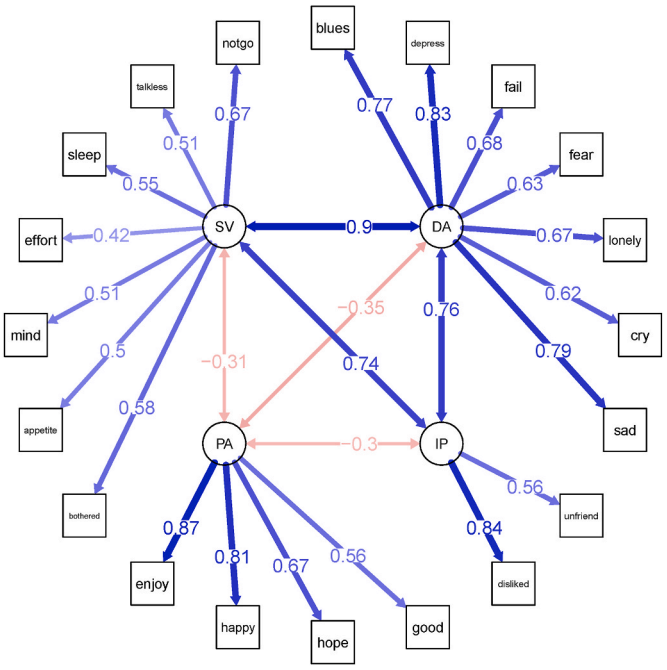


Fig. 1. Factor Structure for both Males and Females
Note. Factor structure was identical for the female subsample ($N = 733$) and for the male subsample ($N = 275$), except for the intercept of the "CRY" item (not shown). All factor loadings and latent inter-correlations were statistically significant (all $ps < .001$).

4. Discussion

In summary, the current work examined six alternative factor structures of the widely-used CES-D-20 using data drawn from 1008 young adults in Singapore. We found that the four-factor structure proposed by the original developer of the scale, Radloff (1977), was the most appropriate structure for the current sample. In addition, we tested measurement invariance across sex, and found that the measure was able to achieve strict invariance with only one intercept (the "CRY" item) released. Moreover, we found equality of latent means and equality of latent variances and covariances across the two sexes.

Our current findings are broadly in line with existing work arguing for the retention of Radloff's (1977) original four-factor solution (e.g., Devins et al., 1988; Hertzog et al., 1990; Nguyen et al., 2004; Williams et al., 2007), providing the first relatively large-scale examination of the standard version of the CES-D-20 in a non-clinical sample from Singapore. Of note, similar efforts have been made to validate the shortened version of the CES-D (i.e., the CES-D-10) in Singapore, such as those by Lee and Chokkanathan (2008) and more recently by Mohebbi et al. (2018). However, the 10-item and the 20-item measures are not necessarily comparable; for example, the 10-item measure has been found to only capture one or two factors, instead of the full four-factor structure. As such, if more nuanced probing of the separate factors is required, then the 20-item version may be more appropriate for use in data collection. While there has been some work examining the CES-D-20 in Singapore, they have mostly neglected to consider its multidimensionality (e.g., Stahl et al., 2008) or have replicated the four-factor structure but only in restricted samples (e.g., caregivers of those with dementia; Ying et al., 2019). Even studies which include more general samples drawn from Singapore (e.g., Kato, 2021) lacked examination of measurement invariance across sex, despite research indicating potential differences in the manifestation of depression between sexes due to variations in hormonal fluctuations, socialisation processes, and coping strategies employed when confronted with stress and depression (Carle et al., 2008; Girgus and Yang, 2015; Hyde and Mezulis, 2020).

The current findings that strict invariance was achieved (with the release of the intercept of the "CRY" item) when comparing the female subsample and the male subsample has some implications for both research and practice. Notably, attaining strict invariance implies that the measure is nearly identical in function for both sexes, despite existing theories positing overall sex differences in the manifestation of depression (e.g., Barbosa-Leiker et al., 2021). Of note, one intercept (of the "CRY" item) was not found to be equal across the two sexes; our findings suggest that for a male individual and female individual who are otherwise identical (e.g., same levels of latent depressive affect), the female individual will tend to have a higher level on the "CRY" item than will her male counterpart. This may be due to cultural expectations and socialisation processes which may differentially affect the expression of crying specifically, which is socially more acceptable for females (Brody, 1993; De Fruyt, 1997; Lombardo et al., 2001; Peter et al., 2001; Santiago-Menendez and Campbell, 2013). Further research should seek to replicate these sex differences (or lack thereof) in the factor structure, both within Singapore, and across other samples.

The current study has several important strengths. Firstly, by utilising a large sample size of over a thousand young adults from Singapore, our research offers significant statistical power and enhances the generalisability of our findings within this demographic. Additionally, our study is among the first to rigorously test the measurement invariance of the CES-D-20 across sexes in Singapore, addressing a critical gap in the existing literature on cultural and gender differences in depression assessment. Utilising this approach not only allowed us to confirm the relevance of the four-factor structure proposed by Radloff (1977), but also illuminate the nuanced ways in which gender influences the expression and reporting of depressive symptoms.

While our study provides important insights into the factor structure of the CES-D-20 in a Singaporean young adult population, it is not without limitations. Primarily, our focus was restricted to young adults, which may limit the applicability of our findings to other age groups. Research by others such as Miller et al. (1997) has suggested that the factor structure of the CES-D-20 can vary significantly with age, noting that older adults might exhibit a two-factor rather than a four-factor structure, irrespective of sex. This indicates potential variations in how depressive symptoms are experienced and reported across different life stages. Consequently, future studies should consider including a broader age range to determine if these findings hold true in more diverse age groups, including children and the elderly, to develop a more comprehensive understanding of the scale's applicability and accuracy.

Table 5
Estimates of internal consistency and average variance extracted.

Factor	Tau-equivalent reliability, α	Congeneric reliability, ω	Average variance extracted, AVE
Depressive affect (7 items)	.891 (male), .872 (female)	.891 (male), .887 (female)	.550 (male), .512 (female)
Positive affect (4 items)	.839 (male), .807 (female)	.774 (male), .841 (female)	.535 (male), .546 (female)
Interpersonal (2 items)	.683 (male), .614 (female)	.664 (male), .708 (female)	.552 (male), .559 (female)
Somatic (7 items)	.779 (male), .720 (female)	.658 (male), .764 (female)	.288 (male), .291 (female)

Note. $N = 1008$.

5. Conclusion

Overall, results from the current study reaffirms the four-factor structure of the CES-D-20 within a Singaporean young adult population, while also revealing generally no sex differences in the expression of depressive symptoms. By validating the four-factor structure and demonstrating strict invariance (with the release of one intercept), we provide robust evidence supporting the CES-D-20's applicability in multicultural settings, while acknowledging subtle but important differences in how depressive symptoms are reported between males and females. These findings underscore the necessity of considering both cultural and gender nuances when employing the CES-D-20 in clinical and research settings, promoting more accurate and culturally sensitive mental health assessments. Future research should aim to explore these factor structures and variances further, particularly across different age groups and cultural backgrounds, to enhance the scale's diagnostic precision and clinical utility worldwide.

CRediT authorship contribution statement

Nadyanna M. Majeed: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. **K.T.A. Sandeeshwara Kasturiratna:** Writing – review & editing, Writing – original draft, Conceptualization. **Nicole R.Y. Chen:** Writing – review & editing, Conceptualization. **Andree Hartanto:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval

All participants provided informed consent. Data collection was approved by the Institutional Review Board at Singapore Management University: IRB-20-118-A075-M3(121), IRB-20-118-A075-M4(621), IRB-20-118-A075-M6(722), and IRB-20-118-A075(920).

Data availability statement

Relevant data and analytic code will be made publicly available on Researchbox #2912 (https://researchbox.org/2912) upon publication.

Funding

Collection of data for this work was supported by research grants from the Ministry of Education Academy Research Fund Tier 1 (20-C242-SMU-001 & 21-SOSS-SMU-023 & 22-SOSS-SMU-041) and Lee Kong Chian Fund for Research Excellence. These grants were awarded to Andree Hartanto by Singapore Management University. The funders had no role in the publication of this manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adler, N.E., Epel, E.S., Castellazzo, G., Ickovics, J.R., 2000. Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. Health Psychol. 19 (6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586.

Barbosa-Leiker, C., Burduli, E., Arias-Losado, R., Muller, C., Noonan, C., Suchy-Dicey, A., Nelson, L., Verney, S.P., Montine, T.J., Buchwald, D., 2021. Gender differences in the assessment of depression in American Indian older adults: the Strong Heart Study. Psychol. Assess. 33 (6), 574–579. https://doi.org/10.1037/pas0001024.

Bentler, P.M., Chou, C.-P., 1992. Some new covariance structure model improvement statistics. Socio. Methods Res. 21 (2), 259–282. https://doi.org/10.1177/0049124192021002006.

Björgvinsson, T., Kertz, S.J., Bigda-Peyton, J.S., McCoy, K.L., Aderka, I.M., 2013. Psychometric properties of the CES-D-10 in a psychiatric sample. Assessment 20 (4), 429–436. https://doi.org/10.1177/1073191113481998.

Blodgett, J.M., Lachance, C.C., Stubbs, B., Co, M., Wu, Y.-T., Prina, M., Tsang, V.W.L., Cosco, T.D., 2021. A systematic review of the latent structure of the Center for Epidemiologic Studies Depression Scale (CES-D) amongst adolescents. BMC Psychiatr. 21 (1), 197. https://doi.org/10.1186/s12888-021-03206-1.

Brody, L.R., 1993. On understanding gender differences in the expression of emotion. In: Human Feelings. Routledge, pp. 87–121.

Cao, C., Liang, X., 2022. The impact of model size on the sensitivity of fit measures in measurement invariance testing. Struct. Equ. Model.: A Multidiscip. J. 29 (5), 744–754. https://doi.org/10.1080/10705511.2022.2056893.

Carle, A.C., Millsap, R.E., Cole, D.A., 2008. Measurement bias across gender on the Children's Depression Inventory: evidence for invariance from two latent variable models. Educ. Psychol. Meas. 68 (2), 281–303. https://doi.org/10.1177/0013164407308471.

Carleton, R.N., Thibodeau, M.A., Teale, M.J.N., Welch, P.G., Abrams, M.P., Robinson, T., Asmundson, G.J.G., 2013. The Center for Epidemiologic Studies Depression scale: a review with a theoretical and empirical examination of item content and factor structure. PLoS One 8 (3), e58067. https://doi.org/10.1371/journal.pone.0058067.

Chakrabarti, A., Ghosh, J.K., 2011. AIC, BIC and recent advances in model selection. In: Philosophy of Statistics. Elsevier, pp. 583–605. https://doi.org/10.1016/B978-0-444-51862-0.50018-6.

Chen, N.R.Y., Majeed, N.M., Goh, A.Y.H., Koh, P.S., Chia, J.L., Hartanto, A., 2024. Dispositional optimism, sleep, and trait affective mediators: a latent variable approach. Pers. Indiv. Differ. 230, 112801. https://doi.org/10.1016/j.paid.2024.112801.

Cheung, G.W., Cooper-Thomas, H.D., Lau, R.S., Wang, L.C., 2024. Reporting reliability, convergent and discriminant validity with structural equation modeling: a review and best-practice recommendations. Asia Pac. J. Manag. 41 (2), 745–783. https://doi.org/10.1007/s10490-023-09871-y.

Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct. Equ. Model.: A Multidiscip. J. 9 (2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.

Cho, E., 2016. Making reliability reliable: a systematic approach to reliability coefficients. Organ. Res. Methods 19 (4), 651–682. https://doi.org/10.1177/1094428116656239.

Cosco, T.D., Prina, M., Stubbs, B., Wu, Y.-T., 2017. Reliability and validity of the Center for Epidemiologic Studies Depression scale in a population-based cohort of middle-aged U. S. Adults. J. Nurs. Meas. 25 (3), 476–485. https://doi.org/10.1891/1061-3749.25.3.476.

De Fruyt, F., 1997. Gender and individual differences in adult crying. Pers. Indiv. Differ. 22 (6), 937–940. https://doi.org/10.1016/S0191-8869(96)00264-4.

Devins, G.M., Orme, C.M., Costello, C.G., Binik, Y.M., Frizzell, B., Stam, H.J., Pullin, W. M., 1988. Measuring depressive symptoms in illness populations: psychometric properties of the Center for Epidemiologic Studies Depression (CES-D) scale. Psychol. Health 2 (2), 139–156. https://doi.org/10.1080/08870448808400349.

Epskamp, S., Cramer, A.O.J., Waldorp, L.J., Schmittmann, V.D., Borsboom, D., 2012. qgraph: network visualizations of relationships in psychometric data. J. Stat. Software 48 (4). https://doi.org/10.18637/jss.v048.i04.

Fifield, J., Reisine, S., 1992. Characterizing the meaning of psychological distress in rheumatoid arthritis. Arthritis care and research : the official journal of the Arthritis Health Professions Association 5 (3), 184–191. https://doi.org/10.1002/art.1790050311.

Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. J. Market. Res. 18 (1), 39. https://doi.org/10.2307/3151312.

Girgus, J.S., Yang, K., 2015. Gender and depression. Current Opinion in Psychology 4, 53–60. https://doi.org/10.1016/j.copsyc.2015.01.019.

Gomez, R., McLaren, S., 2015. The center for epidemiologic studies depression scale: support for a bifactor model with a dominant general factor and a specific factor for positive affect. Assessment 22 (3), 351–360. https://doi.org/10.1177/1073191114545357.

González, P., Nuñez, A., Merz, E., Brintz, C., Weitzman, O., Navas, E.L., Camacho, A., Buelna, C., Penedo, F.J., Wassertheil-Smoller, S., Perreira, K., Isasi, C.R., Choca, J., Talavera, G.A., Gallo, L.C., 2017. Measurement properties of the center for epidemiologic studies depression scale (CES-D 10): findings from HCHS/SOL. Psychol. Assess. 29 (4), 372–381. https://doi.org/10.1037/pas0000330.

Guarnaccia, P.J., Angel, R., Worobey, J.L., 1989. The factor structure of the CES-D in the Hispanic Health and Nutrition Examination Survey: the influences of ethnicity, gender and language. Social Sci. Med. 29 (1), 85–94. https://doi.org/10.1016/0277-9536(89)90131-7.

Hertzog, C., Van Alstine, J., Usala, P.D., Hultsch, D.F., Dixon, R., 1990. Measurement properties of the center for epidemiological studies depression scale (CES-D) in older populations. Psychol. Assess.: J. Consult. Clin. Psychol. 2 (1), 64–72. https://doi.org/10.1037/1040-3590.2.1.64.

Huang, P.-H., 2017. Asymptotics of AIC, BIC, and RMSEA for model selection in structural equation modeling. Psychometrika 82 (2), 407–426. https://doi.org/10.1007/s11336-017-9572-y.

Hurley, A.E., Scandura, T.A., Schriesheim, C.A., Brannick, M.T., Seers, A., Vandenberg, R.J., Williams, L.J., 1997. Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives. J. Organ. Behav. 18 (6), 667–683. https://doi.org/10.1002/(SICI)1099-1379(199711)18:6<667::AID-JOB874>3.0.CO;2-T.

Hyde, J.S., Mezulis, A.H., 2020. Gender differences in depression: biological, affective, cognitive, and sociocultural factors. Harv. Rev. Psychiatr. 28 (1), 4–13. https://doi.org/10.1097/HRP.0000000000000230.

Jorgensen, T.D., Pornprasertmanit, S., Schoemann, A.M., Rosseel, Y., 2022. semTools: useful tools for structural equation modeling, Version 0.5-6). https://CRAN.R-project.org/package=semTools.

Kato, T., 2021. Measurement invariance in the center for epidemiologic studies-depression (CES-D) scale among English-speaking whites and asians. Int. J. Environ. Res. Publ. Health 18 (10), 5298. https://doi.org/10.3390/ijerph18105298.

Kelloway, E.K., 1995. Structural equation modelling in perspective. J. Organ. Behav. 16 (3), 215–224. https://doi.org/10.1002/job.4030160304.

Kim, G., DeCoster, J., Huang, C.-H., Chiriboga, D.A., 2011. Race/ethnicity and the factor structure of the Center for Epidemiologic Studies Depression scale: a meta-analysis. Cult. Divers Ethnic Minor. Psychol. 17 (4), 381–396. https://doi.org/10.1037/a0025434.

Lee, A.E.Y., Chokkanathan, S., 2008. Factor structure of the 10-item CES-D scale among community dwelling older adults in Singapore. Int. J. Geriatr. Psychiatr. 23 (6), 592–597. https://doi.org/10.1002/gps.1944.

Lin, L.-C., Huang, P.-H., Weng, L.-J., 2017. Selecting path models in SEM: a comparison of model selection criteria. Struct. Equ. Model.: A Multidiscip. J. 24 (6), 855–869. https://doi.org/10.1080/10705511.2017.1363652.

Lishinski, 2024. lavaanPlot. Comput. Software, Version 0.8.1. https://CRAN.R-project.org/package=lavaanPlot.

Lombardo, W.K., Cretser, G.A., Roesch, S.C., 2001. For crying out loud—the differences persist into the '90s. Sex. Roles 45 (7/8), 529–547. https://doi.org/10.1023/A:1014862714833.

Lu, Z.-H., Chow, S.-M., Loken, E., 2017. A comparison of Bayesian and frequentist model selection methods for factor analysis models. Psychol. Methods 22 (2), 361–381. https://doi.org/10.1037/met0000145.

Meade, A.W., Johnson, E.C., Braddy, P.W., 2008. Power and sensitivity of alternative fit indices in tests of measurement invariance. J. Appl. Psychol. 93 (3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568.

Miller, T.Q., Markides, K.S., Black, S.A., 1997. The factor structure of the CES-D in two surveys of elderly Mexican Americans. The journals of gerontology. Series B, Psychological Sciences and Social Sciences 52 (5), S259–S269. https://doi.org/10.1093/geronb/52b.5.s259.

Mitchell, A.J., Coyne, J.C., 2007. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br. J. Gen. Pract.: J. Roy. Coll. Gen. Pract. 57 (535), 144–151.

Mohebbi, M., Nguyen, V., McNeil, J.J., Woods, R.L., Nelson, M.R., Shah, R.C., Storey, E., Murray, A.M., Reid, C.M., Kirpach, B., Wolfe, R., Lockery, J.E., Berk, M., 2018. Psychometric properties of a short form of the Center for Epidemiologic Studies Depression (CES-D-10) scale for screening depressive symptoms in healthy community dwelling older adults. Gen. Hosp. Psychiatr. 51, 118–125. https://doi.org/10.1016/j.genhosppsych.2017.08.002.

Nguyen, H.T., Kitner-Triolo, M., Evans, M.K., Zonderman, A.B., 2004. Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. Psychiatr. Res. 126 (2), 177–187. https://doi.org/10.1016/j.psychres.2004.02.004.

Peter, M., Vingerhoets, A.J.J.M., Van Heck, G.L., 2001. Personality, gender, and crying. Eur. J. Pers. 15 (1), 19–28. https://doi.org/10.1002/per.386.

R Core Team, 2023. *R* (version 4.3.1) [Computer software]. https://www.R-project.org/.

Radloff, L.S., 1977. The CES-D Scale: A self-report depression scale for research in the general population. Appl. Psychol. Measure. 1 (3), 385–401. https://doi.org/10.1177/014662167700100306.

Ramírez-Vélez, R., Olabarrieta-Landa, L., Rivera, D., Izquierdo, M., 2023. Factor structure of the 10-item CES-D Scale among patients with persistent COVID-19. J. Med. Virol. 95 (1), e28236. https://doi.org/10.1002/jmv.28236.

Rivera-Medina, C.L., Caraballo, J.N., Rodríguez-Cordero, E.R., Bernal, G., Dávila-Marrero, E., 2010. Factor structure of the CES-D and measurement invariance across gender for low-income Puerto Ricans in a probability sample. J. Consult. Clin. Psychol. 78 (3), 398–408. https://doi.org/10.1037/a0019054.

Rosseel, Y., 2012. lavaan: an R package for structural equation modeling. J. Stat. Software 48 (2). https://doi.org/10.18637/jss.v048.i02.

Rudner, L.M., Schafer, W.D., 2001. *Reliability*. ERIC Clearinghouse on Assessment and Evaluation.

Santiago-Menendez, M., Campbell, A., 2013. Sadness and anger: boys, girls, and crying in adolescence. Psychol. Men Masc. 14 (4), 400–410. https://doi.org/10.1037/a0030661.

Santor, D.A., Gregus, M., Welch, A., 2006. Focus article: eight decades of measurement in depression. Measurement: Interdisciplinary Research & Perspective 4 (3), 135–155. https://doi.org/10.1207/s15366359mea0403_1.

Schroevers, M.J., Sanderman, R., Van Sonderen, E., Ranchor, A.V., 2000. The evaluation of the Center for Epidemiologic Studies Depression (CES-D) scale: depressed and positive affect in cancer patients and healthy reference subjects. Qual. Life Res. 9 (9), 1015–1029. https://doi.org/10.1023/A:1016673003237.

Sheehan, T.J., Fifield, J., Reisine, S., Tennen, H., 1995. The measurement structure of the center for epidemiologic studies depression scale. J. Pers. Assess. 64 (3), 507–521. https://doi.org/10.1207/s15327752jpa6403_9.

Stahl, D., Sum, C.F., Lum, S.S., Liow, P.H., Chan, Y.H., Verma, S., Chua, H.C., Chong, S. A., 2008. Screening for depressive symptoms. Diabetes Care 31 (6), 1118–1119. https://doi.org/10.2337/dc07-2019.

Tavakol, M., Dennick, R., 2011. Making sense of Cronbach's alpha. Int. J. Med. Educ. 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd.

Thorson, J.A., Powell, F.C., 1993. The CES-D: four or five factors? Bull. Psychonomic Soc. 31 (6), 577–578. https://doi.org/10.3758/BF03337359.

Vilagut, G., Forero, C.G., Barbaglia, G., Alonso, J., 2016. Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): a systematic review with meta-analysis. PLoS One 11 (5), e0155431. https://doi.org/10.1371/journal.pone.0155431.

Vrieze, S.I., 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychol. Methods 17 (2), 228–243. https://doi.org/10.1037/a0027127.

Wang, M., Armour, C., Wu, Y., Ren, F., Zhu, X., Yao, S., 2013. Factor structure of the CES-D and measurement invariance across gender in Mainland Chinese adolescents. J. Clin. Psychol. 69 (9), 966–979. https://doi.org/10.1002/jclp.21978.

Wickham, H., 2016. Ggplot2. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4.

Williams, C.D., Taylor, T.R., Makambi, K., Harrell, J., Palmer, J.R., Rosenberg, L., Adams-Campbell, L.L., 2007. CES-D four-factor structure is confirmed, but not invariant, in a large cohort of African American women. Psychiatr. Res. 150 (2), 173–180. https://doi.org/10.1016/j.psychres.2006.02.007.

Williams, L.J., 1995. Covariance structure modeling in organizational research: problems with the method versus applications of the method. J. Organ. Behav. 16 (3), 225–233. https://doi.org/10.1002/job.4030160305.

Wu, P.-C., 2010. Measurement invariance and latent mean differences of the Beck Depression Inventory II across gender groups. J. Psychoeduc. Assess. 28 (6), 551–563. https://doi.org/10.1177/0734282909360772.

Ying, J., Yap, P., Gandhi, M., Liew, T.M., 2019. Validity and utility of the Center for Epidemiological Studies Depression scale for detecting depression in family caregivers of persons with dementia. Dement. Geriatr. Cognit. Disord. 47 (4–6), 323–334. https://doi.org/10.1159/000500940.