Preventive Medicine Reports 24 (2021) 101585

Contents lists available at ScienceDirect

Preventive Medicine Reports

journal homepage: www.elsevier.com/locate/pmedr

Reliability analysis of the Malay version of the center for epidemiologic studies-depression scale (CESD) among adolescents in Malaysia

Sarbhan Singh [a,*], Rafdzah Ahmad Zaki [b], Nik Daliana Nik Farid [b], Kushilpal Kaur [c]

[a] Institute for Medical Research (IMR), National Institutes of Health (NIH), Ministry of Health Malaysia, Setia Alam 40170, Malaysia
[b] Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia
[c] Department of Psychiatry, Hospital Selayang, Ministry of Health, Malaysia, Selangor 68100, Malaysia

ARTICLE INFO

Keywords:
Malay
CESD
Depression
Adolescent
Reliability

ABSTRACT

Depression is a common mental disorder that affects many adolescents worldwide. Therefore, there is a need for reliable instruments to screen for depression symptoms among adolescents.

This study aims to determine the reliability of the Malay version of the Centre of Epidemiological Studies Depression Scale (CESD) among adolescents in Malaysia. A cross-sectional study was conducted among 65 adolescents ages between 12 and 14 years from two secondary schools in the Federal Territory of Kuala Lumpur from May 2017 to July 2017. Cronbach's alpha (α), McDonald's omega (ω), Spearman Brown split half reliability (r_{SB}), and Intra-class Correlation Coefficient (ICC) were examine to determine the internal consistency and two week test-retest reliability. The overall CESD scale was found to have good internal consistency with α = 0.882 (95% CI 0.837, 0.914), ω = 0.886 (95% CI 0.837, 0.916) and r_{SB} = 0.909. The CESD subscales, Somatic symptoms (α = 0.824; 95% CI 0.739, 0.878; ω = 0.828; 95% CI 0.738, 0.885; r_{SB} = 0.825), Depressive affect (α = 0.822; 95% CI 0.745, 0.880; ω = 0.834; 95% CI 0.750, 0.884; r_{SB} = 0.847) and Positive affect (α = 0.610; 95% CI 0.326, 0.721; ω = 0.612; 95% CI 0.379, 0.723 and r_{SB} = 0.608) indicated acceptable to good internal consistency. The 2-week test–retest reliability ICC was 0.926 (95% CI 0.851, 0.961) for the total score reliability. The reliability analysis of the Malay version of CESD shows satisfactory α, ω, r_{SB} and ICC values, therefore making it a reliable instrument to screen for depression among adolescents in Malaysia.

1. Introduction

Depression will be the leading cause of secondary disability worldwide by 2030 as predicted by the World Health Organization (WHO, 2011). During the adolescent phase, depression can impair the functional capability of adolescents thus affecting personal development, interpersonal relationship, and the illness may even persist into adulthood (Latiffah et al., 2016). Depression commonly emerges during mid-adolescence and recurs every 5 to 7 years in 80% of individuals (McCarthy et al., 2011). By 18 years of age, approximately 20% of adolescents will experience an episode of depressive symptoms (McCarthy et al., 2011). The prevalence of adolescent depression is increasing across both Western and Asian countries, wherein the prevalence of depression among adolescents in Malaysia, Thailand, and Indonesia was reported at 29%, 34%, 52.7% during 2013 to 2016 respectively (McCarthy et al., 2011; Mukhripah, 2016; Institute for Public Health. National Health and Morbidity Survey, 2015; Somrongthong et al., 2013). These figures were higher than the previous year's indicating an alarming concern for this problem.

Therefore, it is important to be able to detect symptoms of depression through regular and effective screening using valid and reliable instruments earlier. The commonly used instruments with establish psychometric properties to screen for depression in the general population include Beck Depression Inventory (BDI), Depression, Anxiety and Stress Scale (DASS), Patient Health Questionnaire 9 (PHQ-9) and Children's Depression Inventory (CDI) (Guan, 2014). In addition, the Centre of Epidemiological Study for Depression scale (CESD) is another well establish instrument that is used commonly to screen for depression symptoms. Its original English version was designed in 1977 by the American National Institute of Mental Health and reported to have satisfactory psychometric properties (Radloff, 1977). The CESD instrument has several strengths to it, first it is a systematic tool for screening of depression symptoms as it is comprised of six scales representing various key major domains of depression such as depressed mood,

* Corresponding author.
E-mail address: lssarbhan@moh.gov.my (S. Singh).

https://doi.org/10.1016/j.pmedr.2021.101585
Received 14 April 2021; Received in revised form 23 August 2021; Accepted 3 October 2021
Available online 7 October 2021

feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite and sleep disturbance (Radloff, 1977). In addition, the CESD items were selected from a pool of items from previously validated depression scales such as BDI, CDI and DASS (Radloff, 1977). Due to these reasons, the CESD instrument was selected as the focus of this study to determine its reliability.

Determining the reliability of an instrument ideally should be asses by examining both the internal consistency and test-retest reliability (Cho and Kim, 2015). Internal consistency assesses the correlation between multiple items in a test that are intended to measure the same construct. Several coefficients have been proposed to examine Internal consistency such as Cronbach alpha (α), McDonald's omega (ω) and Spearman Brown split half reliability, however α is widely used due to its familiarity among many researchers and ease of estimation (Cho and Kim, 2015). Nevertheless, studies have recommended the use of several coefficients when examining internal consistency of an instrument (Hayes and Coutts, 2020). The test–retest reliability analysis examines the variation in measurements taken by an instrument on the same subject under the same conditions at different intervals (Ghazali et al., 2016). Selection of an appropriate measure to assess test–retest reliability is important as it should be able to determine both the degree of correlation and agreement between measurement (Chechi and Chakraborty, 2020). The Intra-class Correlation Coefficient (ICC) is considered to be an appropriate measure for test–retest reliability as it reports on correlation and agreement. The ICC is suitable in assessing the reliability of ordinal data such as the Likert Scale as it is a mathematical equivalent of weighted kappa for ordinal data and it addresses limitations of using weighted kappa such as bias and non-independence of ratings (Dekker et al., 2007).

The CESD instrument has been translated to many languages including the Malay language (Mazlan and Ahmad, 2013; Ghazali et al., 2016). The Malay language is a language commonly spoken in many countries across the South East Asia region such as Malaysia; Singapore, Brunei, Indonesia, Thailand, and Southern Philippines. The Malay version of the CESD instrument has been validated previously among the Malaysian general population and found to have satisfactory psychometric properties (Mazlan and Ahmad, 2013; Ghazali et al., 2016). However, previous studies examining the internal consistency reliability of this instrument only reported on the α coefficient without considering other coefficients of internal consistency (Mazlan and Ahmad, 2013; Ghazali et al., 2016). Evidence in literature have reported that the reliance on α as a sole index of reliability is no longer sufficiently warranted due to the existing limitations of α (Hayes and Coutts, 2020). Furthermore, the assumptions of α were not tested prior to its use when assessing the internal consistency reliability of this instrument (Mazlan and Ahmad, 2013; Ghazali et al., 2016). It is important to ensure that assumptions such as unidimensional, Tau-equivalence and homogeneity of items are met prior to the use of α as a measure of internal consistency reliability (9,13). In addition, to date there are no studies that have examined the test–retest reliability of this instrument among Malaysian adolescent. Therefore this study examines the internal consistency and a two-week test–retest reliability of the Malay version CESD instrument among adolescents in Malaysia, by estimating Cronbach alpha, McDonald's omega, Spearman Brown coefficients and ICC respectively.

2. Methods

2.1. Study design and participants

This is a cross-sectional study design that was conducted from May 2017 to July 2017. Participants were students recruited from Secondary Schools in the Federal Territory of Kuala Lumpur. The sampling method was divided into two stages; first two schools were randomly selected from the list of all the secondary schools in Kuala Lumpur (n = 89) which was obtained from the Ministry of Education Malaysia. Second, at the school level, all the students were sampled universally. The inclusion criteria were participants being ages 12 to 14 years with the ability to read and understand the Malay language. The justification to study adolescents aged between 12 and 14 years is because depressive disorders tend to start as early as 12 to 14 year, with its severity increasing across both genders by the age of 12 (Dekker et al., 2007). With the mean age for adolescent depression being 15 and symptoms usually developing two to three years before diagnosis, therefore it is important to screen them early for depression when symptoms are present (Bostic et al., 2012). In addition, studies in Malaysia have reported that young adolescents aged 12 to 13 are significantly more depressed than older adolescents (Latiffah et al., 2016).

Sample size calculation for the reliability analysis was estimated using the Power Analysis and Sample Size (PASS) software version 11.0.7 (NCSS. Power Analysis Sample Size., 2020), with the following parameters Alpha = 0.05, Power = 80%, Z (95% Confidence interval) = 1.96, k (number of raters) = 2, $R_0 = 0.0$ (ICC that is pre-specified in the null hypothesis if it is true) and $R_1 = 0.40$ (ICC that is pre-specified in the alternative hypothesis if it is true) (Fig. 1) (Bujang and Baharum, 2017). A minimum sample size of 37 is sufficient to detect a value of 0.40 for the ICC. An additional 20% drop-out rate was set thus sample size was inflated to 45 participants. Parental consent forms were distributed to 613 students who met the inclusion criteria. A final sample of 65 participants consented to the study and was included for the reliability analysis. This study was approved and met the University of Malaya Research Ethical Committee (UM.TNC 2/UMREC) guidelines for protection of human subjects concerning their safety and privacy

2.2. Data collection

Data collection was conducted at each respective school at a date that was given by the school. Before data collection students were assembled and were briefed on the purpose of this study and their rights as respondents by the researcher. Following which students were given a research information sheet and consent form which the students had to read and bring home to give their guardians or parents to read and sign if agree to participate. Students were advised to return the consent forms to the respective school counselors within 1 to 2 days. After 3 days the researcher receives feedback from the respective school counselor on the number of consent forms returned. A subsequent date was set whereby the researcher distributes the Malay version CESD questionnaire to the participant in the respective schools. All questionnaires were coded and

$$n = 1 + \frac{2(Z_\alpha + Z_\beta)^2 k}{(\ln C_0)^2 (k-1)}$$

$$\text{where: } C_0 = \frac{1 + k\Theta_0}{1 + k\Theta_1}$$

$$\Theta_0 = \frac{R_0}{1 - R_0}$$

$$\Theta_1 = \frac{R_1}{1 - R_1}$$

Fig. 1. Reliability analysis (ICC) sample size calculation.

have no personal identifiers written on them. This is to ensure the confidentiality of participants was maintained. The average time taken to complete the questionnaire was 10 min. After the session, the researcher collected all the questionnaires from the participants. To minimize missing data firstly the researcher advised all participants to check their questionnaire at the end to ensure all questions have been answered and secondly the researcher checks all the returned questionnaires immediately upon collection to detect any missing data in the school itself. Test-retest reliability was conducted within a two week interval at the respective schools. The same procedures as above were repeated for the distribution and collection of questionnaires. The study flow is shown in Fig. 2.

2.3. Measures

The Malay version of the Center for Epidemiology Study Depression Scale (CESD) is a 20-item self-administered questionnaire based on a 4-point Likert scale response options which are 0 (none of the time) to 3 (most all of the time). Item numbers 4,8,12 and 16 are scored in the reverse order. The total score for all items will provide a continuous value of CESD score from 0 to a maximum of 60. A cut-off point of 27 has been selected to indicate depression symptoms among adolescents in Malaysia (Ghazali et al., 2016). Respondents were then classified as having depression or non-depression symptoms. Since this tool is available in the Malay language no translation was required. Permission to use the Malay version Center for Epidemiology Study Depression Scale was obtained from the respective authors (Ghazali et al., 2016).

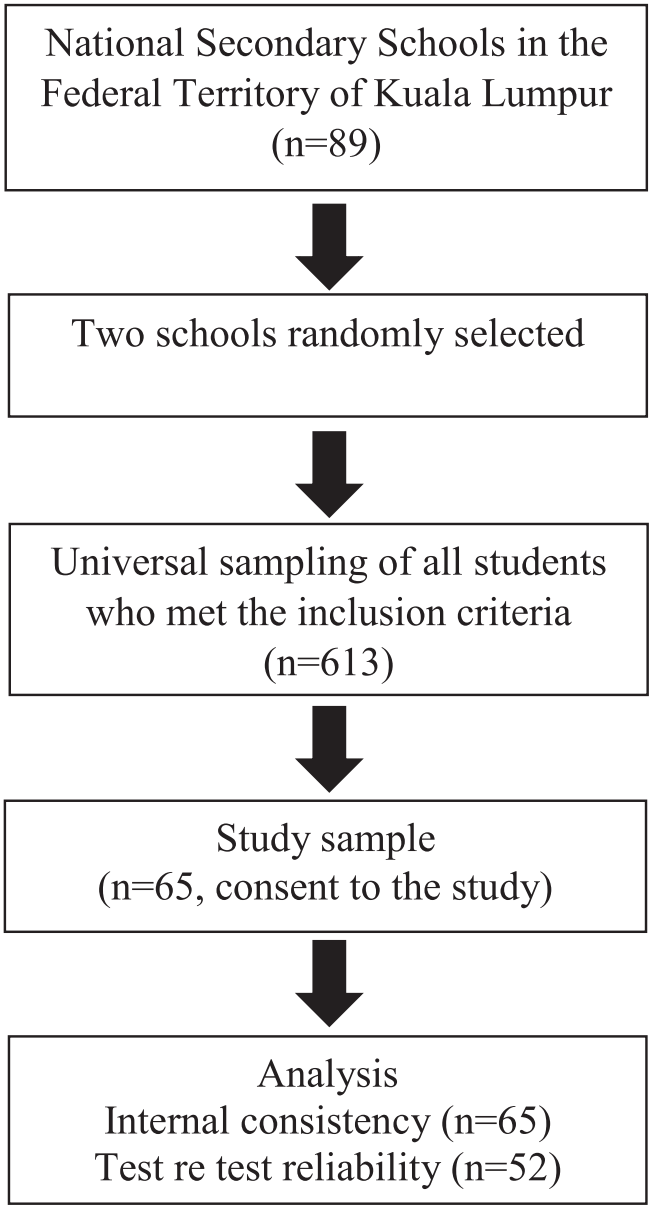

Fig. 2. Study Flow.

2.4. Analysis

Data were analyzed using Statistical Program for the Social Sciences (SPSS) version 24.0 and R programming software version 4.1.1. Double data entry was performed to ensure accuracy of data and minimization of error, data were checked for missing data and abnormal values before performing any statistical analysis. There were no missing or abnormal values. The descriptive analysis included frequency of socio-demographic characteristics, depression scores (CESD), and percentages of floor and ceiling effect. The threshold for a significant floor or ceiling effect was set at 15% (Terwee et al., 2007).

To assess the internal reliability of the Malay version of the CESD scale, Cronbach alpha (α), McDonald's Omega coefficient (ω) and Spearman Brown split half reliability for internal consistency were calculated using the R package Coefficient alpha and SPSS respectively (Hayes and Coutts, 2020; Zhang and Yuan, 2016; Gliem and Gliem, 2003; Eisinga et al., 2013). The reference values of the α and ω are categorized as the following < 0.50, 0.50 to < 0.60, 0.60 to < 0.70, 0.70 to < 0.80, 0.80 to 0.90 and > 0.90 which indicates unacceptable, poor, questionable, acceptable, good, and excellent internal consistency, respectively (Gliem and Gliem, 2003). Spearman Brown coefficient values of 0.80 and above indicates good internal consistency (Eisinga et al., 2013). The difference between the reliability coefficient (α and ω) was examined by estimating the reliability coefficient Confidence Interval (CI) difference wherein if the CI do not contain 0 then it indicates a significant difference between the reliability coefficients (Lifang and Wai, 2016). Prior to conducting the reliability analysis for internal consistency, the assumptions of constructs being unidimensional, Tau-equivalence and homogeneity of items were tested (9,13). Wherein unidimensionality was tested using Exploratory Factor Analysis (EFA) in SPSS and Confirmatory Factor Analysis (CFA) using the R package Laavan (Rosseel, 2012), while Tau-equivalence and homogeneity of items was tested using R package Coefficient alpha (Zhang and Yuan, 2016) respectively. Model fit estimates namely factor loading > 0.5, Chisquare $p > 0.05$, Tucker-Lewis Index (TLI) > 0.9, Root Mean Square Error of Approximation (RMSEA) < 0.08, Goodness of fit (GFI) > 0.9, Adjusted GIF > 0.9, Non fit index (NFI) > 0.9 and Comparative fit index (CFI) > 0.9 supports the assumptions of constructs being unidimensional (Nazim and Ahmad, 2013). While p values < 0.05 indicates non Tau-equivalence and non-homogeneity of items (Zhang and Yuan, 2016); α requires the assumption of unidimensional constructs, Tau-equivalence and homogeneity of items, however in case any of these assumptions are violated then ω should be used as a substitute for alpha (Cho and Kim, 2015).

ICC was estimated in SPSS to evaluate the absolute agreement for the total score and each item. The ICC estimates and their 95% confident intervals were calculated based on absolute-agreement and 2-way mixed-effects model. The reference values of the ICC are categorized as the following < 0.5, 0.5 to < 0.75, 0.75 to < 0.9, and > 0.90 which indicates poor, moderate, good, and excellent reliability, respectively based on 95% CI (Portney and Watkins, 2009).

3. Results

3.1. Participant characteristics

Majority of participant's were females, with ages ranging from 12 to

14 years with a mean of 13.1 years. The prevalence of depression was 32.3% (Table 1). There were no floor and ceiling reported as the percentages of participants having the lowest and highest possible score were zero respectively.

3.2. Test for Unidimensionality, Tau-equivalence and homogeneity of items

The EFA using the Principal Components Analysis and varimax rotation method conducted on CESD items showed that the Kaiser-Meyer-Olkin index of sampling adequacy for CESD was 0.76, which is above the recommended value of 0.60, and the Barlett's test of sphericity for each alternative form was significant for the sample ($\chi 2$ (190) = 535.9, $p < 0.001$). These indicated that the data represented a homogeneous collection of variables that were suitable for factor analysis. Three factors were extracted from the factor analysis with eigen-values larger than one, explaining 49.51% of the total variance (Table 2). Factor 1 accounted for 33.18% of the variance and contained nine items all characterized by somatic symptoms. Factor 2 accounted for 9.19% of the variance and contained seven items all characterized by depressive affect. Factor 3 accounted for 7.12% of the variance and consists of four items that include a general range of expressive items about positive affect. The items loadings on each of the extracted factors are shown in Table 3. Subsequently the three factor model identified from EFA was tested for unidimensionality using CFA, wherein the overall model fit estimates were indicative for unidimensionality with factor loading > 0.50 (Table 4), chi square p = 0.979, TLI = 1.155, RMSEA = 0.002, GFI = 0.960, Adjusted GFI = 0.924, NFI = 0.921 and CFI = 0.999. Similarly, items within the somatic symptom, depressive and positive affect subscales were approximately unidimensional with factor loading > 0.50, chi square p > 0.05, TLI (0.961, 0.961 and 1.091 respectively), RMSEA (0.048, 0.059 and 0.001 respectively), GFI (0.913, 0.930 and 0.991 respectively), Adjusted GFI (0.855, 0.861, 0.956 respectively) and NFI (0.819, 0.887 and 0.963 respectively).

All three constructs were found to be tau equivalent and homogeneous. The Somatic construct reported a F statistic 1.48 with a p-value 0.140 (*Test of tau equivalent*) and a F statistic 1.105 with a p-value 0.383 (*Test of homogeneous items*). The depressive affect construct reported a F statistic 1.825 with a p-value 0.055 (*Test of tau equivalent*) and a F statistic 1.024 with a p-value 0.450 (*Test of homogeneous items*). The positive affect construct reported a F statistic 1.129 with a p-value 0.355 (*Test of tau equivalent*) and a F statistic 0.486 with a p-value 0.617 (*Test of homogeneous items*). With the assumptions of unidimensionality, tau equivalent and homogeneity of items being sufficiently meet, the use of α, ω and Spearman Brown analysis for internal consistency is justified.

Table 1
Summary of participants' demographic information (n = 65).

Demographic information	N (%)
Gender	
Male	13 (20)
Female	52 (80)
Age	
12 yr	1 (1.5)
13 yr	60 (92.3)
14 yr	4 (6.2)
Ethnicity	
Malay	21 (32.3)
Chinese	37 (56.9)
Indian	2 (3.1)
Others	5 (7.7)
Depression*	
Yes	44 (67.7)
No	21 (32.3)

Note. * Prevalence of depression based on overall scores 27 and above.

Table 2
Total variance explained based on Principal component analysis for CESD scale.

Factor	Eigenvalues value	Percentage (%) of Variance	Cumulative Percentage (%)
1	6.637	33.187	33.187
2	1.840	9.198	42.384
3	1.425	7.123	49.508

Table 3
Factor loadings based on Principle Component Analysis with varimax rotation for items that loaded, each question.

Item	Factor loading		
	1	2	3
11. My sleep was restless	0.767	−0.013	0.113
9. I thought my life had been a failure	0.727	0.083	−0.019
18. I felt sad	0.628	0.556	0.007
3. I felt that I could not shake off the blues	0.571	0.138	0.119
5. I had trouble keeping my mind on what I was doing	0.560	0.276	0.064
6. I felt depressed	0.495	0.305	0.301
7. I felt that everything I did was an effort	0.482	0.330	0.377
1. I was bothered by things that usually don't bother me	0.474	0.400	−0.106
15. People were unfriendly	0.462	0.270	−0.130
10. I felt fearful	0.366	0.688	0.267
20. I could not get going	0.355	0.649	−0.033
13. I talked less than usual	0.043	0.614	0.098
14. I felt lonely	0.135	0.614	0.287
17. I had crying spells	0.452	0.612	0.096
2. I did not feel like eating; my appetite was poor	0.194	0.584	−0.386
19. I felt that people dislike me	0.381	0.536	0.312
16. I enjoyed life	−0.104	0.300	0.682
12. I was happy	0.430	0.098	0.671
8. I felt hopeful about the future	0.077	−0.180	0.536
4. I felt I was just as good as other people	−0.063	0.396	0.514

Note. Boldface indicates highest factor loadings.

Table 4
Standardized factor loadings of items based on CFA.

Item	Factor loading		
	1	2	3
11. My sleep was restless	0.752		
9. I thought my life had been a failure	0.612		
18. I felt sad	0.607		
3. I felt that I could not shake off the blues	0.607		
5. I had trouble keeping my mind on what I was doing	0.592		
6. I felt depressed	0.589		
7. I felt that everything I did was an effort	0.544		
1. I was bothered by things that usually don't bother me	0.501		
15. People were unfriendly	0.500		
10. I felt fearful		0.851	
20. I could not get going		0.732	
13. I talked less than usual		0.709	
14. I felt lonely		0.612	
17. I had crying spells		0.609	
2. I did not feel like eating; my appetite was poor		0.529	
19. I felt that people dislike me		0.501	
16. I enjoyed life			0.792
12. I was happy			0.632
8. I felt hopeful about the future			0.557
4. I felt I was just as good as other people			0.524

3.3. Reliability analysis

The Cronbach alpha (α), McDonald's omega (ω), Spearman Brown coefficients (r_{SB}) values were α = 0.882 (95% CI 0.837, 0.914), ω = 0.886 (95% CI 0.837, 0.916) and r_{SB} = 0.909 respectively, indicating good to excellent internal consistency for the overall CESD Malay version scale. For the CESD subscales, the Somatic symptoms construct

(α = 0.824; 95% CI 0.739, 0.878; ω = 0.828; 95% CI 0.738, 0.885; r_{SB} = 0.825) and Depressive affect construct (α = 0.822; 95% CI 0.745, 0.880; ω = 0.834; 95% CI 0.750, 0.884; r_{SB} = 0.847) showed good internal consistencies. The Positive affect construct was found to have acceptable levels of internal consistency with α = 0.610; 95% CI 0.326, 0.721; ω = 0.612; 95% CI 0.379, 0.723 and r_{SB} = 0.608.

Overall the ω and r_{SB} coefficient estimates were consistency higher than α values for the overall and the respective subscales. Where in there is a significant difference (as the CI do not contain 0) between the ω and α reliability estimates for the overall scale (reliability difference ω-α = 0.004, 95%CI 0.001,0.007), Somatic symptom subscale (reliability difference ω-α = 0.004, 95%CI 0.002,0.008), Depressive affect subscale (reliability difference ω-α = 0.012, 95%CI 0.008,0.018) and Positive affect subscale (reliability difference ω-α = 0.002, 95%CI 0.001,0.006). Table 5 shows the Internal consistency reliability analysis for the CESD scale. In addition, Item-wise deletion revealed that α did not increase by>0.009 with the exclusion of any item. The CITC and item-wise deletion α values of all items are shown in Table 6. The test–retest reliability for the total CESD score ICC was 0.926 (95% CI 0.851, 0.961) indicating excellent reliability. For item by item reliability analysis the ICC ranged from the lowest value ICC = 0.499 (95% CI 0.143, 0.710) for item 14 to the highest value ICC = 0.872 (95% CI 0.778, 0.927) for item 18. Overall among the 20 items, 12 items (57.1%), 7 items (33.3%), 1 item (4.8%) showed moderate, good, excellent and poor reliability respectively (Table 7).

4. Discussion

In this study, we examined the Cronbach alpha, McDonald's omega, Spearman Brown coefficients and ICC as measures of internal consistency and test–retest reliability respectively for the Malay version CESD instrument among adolescents in Malaysia. The first part of this study focused on examining the assumptions of Unidimensionality, Tau-equivalence and homogeneity of items which are required to be satisfied prior to the use of internal consistency reliability measures. Wherein all these assumptions were sufficiently met for the CESD instruments and its subscales. Many studies in the past have ignored these assumptions when examining the reliability analysis of the Malay CESD instrument (Mazlan and Ahmad, 2013; Ghazali et al., 2016). This could result in the use of inappropriate reliability measures (i.e α when Unidimensionality is not present or ω for non Tau-equivalence) which potentially will result in underestimation of true α, inconsistent and incomparable reliability estimates (Chechi and Chakraborty, 2020).

Following meeting these assumptions, this study reported good to excellent point estimates of α, ω and Spearman Brown coefficients internal consistency reliability measures with fairly narrow CIs for the overall CESD scale (point estimate reliability range 0.882 to 0.909), and its subscale; Somatic symptoms (point estimate reliability range 0.824 to 0.828) and Depressive affect (point estimate reliability range 0.822 to 0.884). While the Positive affect subscale was found to have acceptable point estimates levels of internal consistency along with wider CIs (point estimate reliability range 0.610 to 0.612). Similarly, findings have been reported previously wherein the Positive affect subscale was found to have lower levels of internal consistencies compared to other subscales (Mazlan and Ahmad, 2013; Ghazali et al., 2016; Cheung et al., 1998; Sharif Nia et al., 2019). This might due to the small number of items in the Positive affect subscale (n = 4) and the proposition that positive affect items should not be part of the general depression factor (Cheung et al., 1998; Mohsen and Dennick, 2011). In addition, the overall α = 0.882 estimate for the CESD scale in this study is higher compared to previous studies looking at the same (α = 0.74 to 0.85) (Mazlan and Ahmad, 2013; Ghazali et al., 2016). Aside from methodological variations among studies, this differences could be attributed to α underestimation when applied in the context of non unidimensionality or non Tau-equivalence (Chechi and Chakraborty, 2020).

Table 6
CITC and Item-wise deletion Cronbach's alpha values.

CESD Item	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item 1	0.471	0.877
Item 2	0.334	0.880
Item 3	0.459	0.877
Item 4	0.334	0.881
Item 5	0.514	0.875
Item 6	0.553	0.875
Item 7	0.609	0.872
Item 8	0.076	0.891
Item 9	0.461	0.877
Item 10	0.728	0.868
Item 11	0.484	0.876
Item 12	0.518	0.875
Item 13	0.403	0.879
Item 14	0.527	0.875
Item 15	0.395	0.879
Item 16	0.281	0.882
Item 17	0.679	0.869
Item 18	0.730	0.868
Item 19	0.647	0.870
Item 20	0.592	0.873

Table 7
Intra Class Correlation and 95% CI for total and individual items.

Items	Intra Class Correlation (ICC)	95% CI
Item 1	0.550	0.220, 0.741
Item 2	0.799	0.649, 0.885
Item 3	0.612	0.319, 0.778
Item 4	0.554	0.233, 0.742
Item 5	0.735	0.541, 0.847
Item 6	0.590	0.282, 0.766
Item 7	0.812	0.673, 0.892
Item 8	0.711	0.496, 0.834
Item 9	0.581	0.267, 0.760
Item 10	0.764	0.574, 0.868
Item 11	0.502	0.151, 0.710
Item 12	0.836	0.715, 0.906
Item 13	0.699	0.479, 0.827
Item 14	0.499	0.143, 0.710
Item 15	0.680	0.443, 0.817
Item 16	0.719	0.509, 0.839
Item 17	0.829	0.697, 0.903
Item 18	0.872	0.778, 0.927
Item 19	0.806	0.663, 0.888
Item 20	0.715	0.501, 0.836
TOTAL	0.926	0.851, 0.961

Table 5
Internal consistency reliability analysis.

	Cronbach's alpha (α)			McDonald's omega (ω)			Reliability Difference	
	Alpha	95%CI	SE	Omega	95%CI	SE	ω-α	95%CI
Overall CESD	0.882	0.837,0.914	0.020	0.886	0.837,0.916	0.019	0.004	0.001,0.007
Somatic symptoms subscale	0.824	0.739,0.878	0.036	0.828	0.738,0.885	0.039	0.004	0.002,0.008
Depressive affect subscale	0.822	0.745,0.880	0.036	0.834	0.750,0.884	0.033	0.012	0.008,0.018
Positive affect subscale	0.610	0.326,0.721	0.102	0.612	0.379,0.723	0.201	0.002	0.001,0.006

Note. SE; Standard Errors.

This study found that the ω and Spearman Brown coefficients point estimates were consistency higher than α values for the overall CESD and its subscales. Furthermore, this difference is statistically significant between ω and α estimates. These findings suggest that despite the difference between ω and α being minor (0.004), it's indicative of a significantly higher ω estimates compared to α. Evidence in literature supports our findings of higher ω and Spearman Brown coefficients estimates compared to α (Lifang and Wai, 2016; Sharif Nia et al., 2019; Roberts, 1980). However, far fewer studies have attempted to test the difference between these coefficients. In addition, this study grossly found fairly consistent categorization of α, ω and Spearman Brown coefficients internal consistency reliability measures for the overall CESD and its subscales. Wherein the estimates were well within the similar category range (i.e acceptable, moderate, good, excellent) for the respective reliability estimates. This increases the overall confidence when reporting the internal consistency level for the CESD instrument instead of just relying on a single coefficient which is no longer sufficient (Hayes and Coutts, 2020).

The test-retest reliability in this study was estimated via the ICC (point estimate and 95% CI) which was estimated using a 2-way mixed-effects model with absolute agreement. This model is an appropriate one as it is indicated in test-retest reliability in which repeated measurements cannot be regarded as randomized samples and measurements would be meaningless if there is no agreement between repeated measurements (Portney and Watkins, 2009). This study reported a total ICC value of 0.926 (95% CI 0.851, 0.961) indicating excellent reliability. Similarly, previous studies have reported total ICC values which are close to that in this study at 0.85 and 0.91 respectively (Miller et al., 2008; Chin et al., 2015). The could be due to the almost similar sample size used for test-retest reliablity analysis which ranged from 47 to 52 subjects and the similar 2-week re-test interval duration used.

Item wise ICC analysis indicated that majority of the items were categorized to have moderate (57.1%) to good (33.3%) reliability. Among the items that were classified to have poor to moderate reliability, many were from the somatic symptom construct. Difficultly in understanding the somatic symptoms of depression among young adolescents and subject variability factors could possibly result in this findings (van Beljouw et al., 2010). Several approaches could be taken to improve the reliability of these items in future studies. Among them are using a larger study sample, rewording the items and increasing the awareness of somatic symptoms among adolescents. It is important for young adolescents to be aware of the somatic symptoms of depression in addition to the depressive affect. In addition, the lowest and highest ICC values reported in this study were for item 14 (ICC = 0.49) and item 18 (ICC = 0.87) respectively. Studies done by Miller et al (Miller et al., 2008) and Tatar et al (2010) had reported the lowest correlation values at 0.33 (item 4) and 0.09 (item 11), and the highest correlation values at 0.74 (item 3) and 0.57 (item 10) respectively (Miller et al., 2008; Tatar et al., 2013). This difference could be due to methodological variations (i.e 2-way mixed-effects model or 2-way random model) in estimating ICC (Koo and Li, 2016).

Among the strengths of this study include ensuring that the assumptions of Unidimensionality, Tau-equivalence and homogeneity of items are satisfied prior to the application of internal consistency reliability measures. Also this study provides point estimates and CI of several coefficients (α, ω and Spearman Brown) to examine the internal consistency of the CESD instrument. By doing so we are able to indicate the degree of confidence the true estimate would fall within the 95% CI range and also address limitations associated with the use of a single coefficient and are able to uniformly conclude on the internal consistency reliability (Hayes and Coutts, 2020). In addition, a test re-test was performed within a 2-week interval which would likely reduce recall bias at the same time reducing the potential for change in symptoms of depression (Streiner and Norman, 2003). Finally, this is the first study to examine the internal consistency (using several coefficients) and the test–retest reliability of the Malay version of the CESD instrument among young adolescent population in Malaysia. Limitations of the current study include that majority of the participants were 13 years of age, this would limit the generalizability of our study results to those older adolescents, however, our findings could be generalized to adolescent within the early adolescent age range which is between 10 and 14 years of age (Sawyer et al., 2012). The dropout rate reported by our study is 20%, wherein 52 out of 65 participants were present for the second administration of the questionnaire (T2), however, this was accounted for during the sample size calculation.

5. Conclusion

The overall reliability analysis of the Malay version of CESD shows satisfactory Cronbach alpha, McDonald's omega, Spearman Brown and ICC values, therefore making it a reliable instrument to screen for depression among adolescents in Malaysia.

CRediT authorship contribution statement

Sarbhan Singh: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Rafdzah Ahmad Zaki:** Conceptualization, Methodology, Software, Validation, Writing – review & editing, Supervision. **Nik Daliana Nik Farid:** Conceptualization, Methodology, Software, Validation, Writing – review & editing, Supervision. **Kushilpal Kaur:** Writing – original draft, Writing – review & editing, Visualization.

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.

Acknowledgments

We would like to thank the Director General of Health Malaysia for his permission to publish this article. Also we would like to thank all the students who participated in this study and gave full cooperation during data collection. We are also very grateful to Ministry of Education, Malaysia for greatly assisting the research process.

Funding

No financial support was taken in this study.

Competing interest

The authors have no conflicts of interest, no competing financial interests and personal relationships to declare that could have appeared to influence the work reported in this paper. All authors have read and approved the final version of the manuscript.

Ethical declarations

The study was registered with National Medical Research Register (NMRR-18-719-40569). Ethics approval was obtained from University of Malaya Research Ethical Committee (UM.TNC 2/UMREC). Permission to use the School was approved by the Malaysian Ministry of Education, Kuala Lumpur State Education Department and respective School Principals. Written consent was obtained from the participant's guardians. Permission to use and examine the reliability of the Malay version of the CESD instrument was obtained from the original authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.pmedr.2021.101585.

References

WHO. Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level. 2011. Available from: https://apps.who.int/gb/ebwha/pdf_files/EB130/B130_9-en.pdf.

Latiffah, L., Tajik, E., Ibrahim, N., Abubakar, A.S., Ali, S.S.B., 2016. Depression and its associated factors among secondary school students in Malaysia. Southeast Asian J. Trop. Med. Public Health 47 (1), 131–141.

McCarthy, J., Bruno, M., Fernandes, T.E., 2011. Evaluating mental health literacy and adolescent depression: what do teenagers“ know?”. Prof Couns. 1 (2), 133–142.

Mukhripah, D., 2016. Prevalence and predictors of depression among high school students in Indonesia. J. Nurs. Sci. Heal. 39 (1), 132–145.

Institute for Public Health. National Health and Morbidity Survey 2015: Non-Communicable Diseases, Risk Factors & Other Health Problems . Ministry of Health. Kuala Lumpur, Malaysia; 2015 [cited 2018 Apr 18]. p. 1–291. Available from: http://iku.moh.gov.my/images/IKU/Document/REPORT/nhmsreport2015vol2.pdf.

Somrongthong, R., Wongchalee, S., Laosee, O., 2013. Depression among adolescents: a study in a Bangkok slum community. Scand. J. Caring. Sci. J. Caring. Sci. 27 (2), 327–334.

Guan, N.C., 2014. A review of depression research in Malaysia. Med. J. Malaysia 69 (8), 42–45.

Radloff, L.S., 1977. The CES-D scale: a self-report depression scale for research in the general population. Appl. Psychol. Meas. 1 (3), 385–401.

Cho, E., Kim, S., 2015. Cronbach’s coefficient alpha: well known but poorly understood. Organ. Res. Methods. 18 (2), 207–230.

Hayes, A.F., Coutts, J.J., 2020. Use omega rather than cronbach’s alpha for estimating reliability. But…. Commun. Methods Meas. 14 (1), 1–24.

Mazlan, N.H., Ahmad, A., 2013. Validation of the malay-translated version of the center for epidemiological study–depression scale (CES-D). ASEAN J. Psychiatry. 15 (1), 54–65.

Ghazali, S., Elklit, A., Balang, R., Chen, Y., 2016. Concurrent validity and exploratory factor analysis of the malay version of center for epidemiologic studies-depression scale (CESD) among malaysian adolescents. ASEAN J. Psychiatry. 17 (1), 71–78.

Chechi, V.K., Chakraborty, R., 2020. Testing the assumptions of tauequivalence and reliability analysis in the academic emotion regulation questionnaire (AERQ) subscales. J. Crit. Rev. 7 (16), 1830–1838.

Dekker, M.C., Ferdinand, R.F., van Lang, N.D.J., Bongers, I.L., van der Ende, J., Verhulst, F.C., 2007. Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome. J. Child Psychol. Psych. 48 (7), 657–666.

Bostic, J.Q., Bagnell, A.L. Evidence-Based School Psychiatry, An Issue of Child and Adolescent Psychiatric Clinics of North America . Vol. 2. Halifax, NS, Canada; 2012. 241 p. Available from: https://www.elsevier.com/books/evidence-based-school-psychiatry-an-issue-of-child-and-adolescent-psychiatric-clinics-of-north-america/bostic/978-1-4557-3839-7.

NCSS. Power Analysis & Sample Size. 2020 [cited 2021 Jun 6]. Available from: https://www.ncss.com/software/pass/.

Bujang, M.A., Baharum, N., 2017. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. J. Sch. Dent. Sci. USM. 12 (1), 1–11.

Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.M., de Vet, H.C.W., 2007. Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. 60 (1), 34–42.

Zhang, Z., Yuan, K.-H., 2016. Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: methods and software. Educ. Psychol. Meas. 76 (3), 387–411.

Gliem, J.A., Gliem, R.R. Calculating, interpreting, and reporting Cronbach’s alpha reliability coefficient for Likert-type scales. In: Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education. Ohio State, Columbus; 2003. p. 82–8.

Eisinga, R., Grotenhuis, M.t., Pelzer, B., 2013. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? Int J Public Health. 58 (4), 637–642.

Lifang, D., Wai, C., 2016. Testing the difference between reliability coefficients alpha and omega. Educ. Psychol. Meas. 77 (2), 185–203.

Rosseel, Y., 2012. lavaan: an r package for structural equation modeling. J. Stat. Softw. 48 (2), 1–36.

Nazim, A., Ahmad, S., 2013. Assessing the unidimensionality, reliability, validity and fitness of influential factors of 8th grades student’s Mathematics achievement in Malaysia. Int. J. Adv. Res. 1 (2).

Portney, L.G., Watkins, M.P. Foundations of clinical research: applications to practice. New Jersey: Prentice Hall. 2009.

Cheung, C.K., Bagley, C. Validating an American scale in Hong Kong: the Center for Epidemiological Studies Depression Scale (CES-D). J Psychol. 132AD;2:169–86.

Sharif Nia, H., Rezapour, M., Allen, K.A., Pahlevan Sharif, S., Jafari, A., Torkmandi, H., Goudarzian, A.H., 2019. The Psychometric Properties of the Center for Epidemiological Studies Depression Scale (CES-D) for Iranian Cancer Patients. Asia Pacific. J. Cancer Prev. 20 (9), 2803–2809.

Mohsen, T., Dennick, R., 2011. Making sense of Cronbach’s alpha. Int. J. Med. Educ. 2, 53–55.

Roberts, R.E., 1980. Reliability of the CES-D Scale in different ethnic contexts. Pscyhiatry Res. 2 (2), 125–134.

Miller, W.C., Anton, H.A., Townson, A.F., 2008. Measurement properties of the CESD scale among individuals with spinal cord injury. Spinal Cord. 46 (4), 287–292.

Chin, W.Y., Choi, E.P.H., Chan, K.T.Y., Wong, C.K.H., Chilcot, J., 2015. The psychometric properties of the center for epidemiologic studies depression scale in chinese primary care patients: factor structure, construct validity, reliability, sensitivity and responsiveness. Chilcot J, editor. PLoS One 10 (8), e0135131. https://doi.org/10.1371/journal.pone.0135131.

van Beljouw, I.M., Verhaak, P.F., Cuijpers, P., van Marwijk, H.W., Penninx, B.W., 2010. The course of untreated anxiety and depression, and determinants of poor one-year outcome: a one-year cohort study. BMC Psychiatry. Oct 20 [cited 2017 Apr 27];10:86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20961414.

Tatar, A., Kayiran, S.M., Saltukoglu, G., Ozkut, E.Ş.Z., Emeksiz, M., 2013. Analysis of the Center for Epidemiologic Studies Depression Scale (CES-D) in children and adolescents from the perspective of the item response theory. Bull. Clin. Psychopharmacol. 23 (3), 242–253.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 15 (2), 155–163.

Streiner, D., Norman, G. Pretty Darned Quick (PDQ) statistics. 3rd ed. Shelton C, editor. United States: People’s medical publishing house; 2003.

Sawyer, S.M., Afifi, R.A., Bearinger, L.H., Blakemore, S.-J., Dick, B., Ezeh, A.C., Patton, G.C., 2012. Adolescence: a foundation for future health. Lancet 379 (9826), 1630–1640.