Factor Structure of the CES-D and Measurement Invariance Across Gender in Mainland Chinese Adolescents

Mengcheng Wang,1 Cherie Armour,2 Yan Wu,3 Fen Ren,4 Xiongzhao Zhu,5 and Shuqiao Yao3

1Guangzhou University, Guangzhou, China
2University of Ulster, Coleraine, North Ireland
3Guangdong University of Foreign Studies, Guangzhou, China
4Institute of Psychology, Chinese Academy of Sciences, Beijing, China
5Central South University, Changsha, China

Objectives: The primary aim was to examine the depressive symptom structure of Mainland China adolescents using the Center for Epidemiologic Studies Depression Scale (CES–D). Design: Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were simultaneously conducted to determine the structure of the CES-D in a large scale, representative adolescent samples recruited from Mainland China. Multigroup CFA (N = 5059, 48% boys, mean = 16.55±1.06) was utilized to test the factorial invariance of the depressive symptom structure, which was generated by EFA and confirmed by CFA across gender. Results: The CES-D can be interpreted in terms of 3 symptom dimensions. Additionally, factorial invariance of the new proposed model across gender was supported at all assuming different degrees of invariance. Conclusion: Mainland Chinese adolescents have specific depressive symptom structure, which is consistent across gender. © 2013 Wiley Periodicals, Inc. J. Clin. Psychol. 69:966–979, 2013.

Keywords: depressive symptom; CES–D; factor structure; factorial invariance; Chinese adolescent

The Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977) is a commonly employed self-report measure of depressive symptomatology. The CES-D was originally designed and developed for the identification of depression among the general adult population. The scale items cover the major components of depressive symptomatology and were selected from previously validated depression scales. The CES-D has since been utilized across various age groups, including elderly people (e.g., Zhang et al., 2011), adult community samples (e.g., Roberts, & Vernon, 1983), college students (e.g., Yen, Robins, & Lin, 2000), adolescents (e.g., Cheng, Yen, Ko, & Yen, 2012; Lee et al., 2008; Radloff, 1991; Roberts, Andrews, Lewinsohn, & Hops, 1990), and children (e.g., Li, Chung, & Ho, 2010).

Radloff (1977) originally proposed that the 20 CES-D items were categorized into four symptom groups: depressed affect (DA; seven items); somatic complaints (SC; seven items); interpersonal problems (IP; two items); and positive affect (PA; four items). Subsequent support for the four-factor model has been plentiful across various populations from the United States (e.g., Golding, & Aneshensel, 1989; Hertzog, Van Alstine, Usala, Hultsch, & Dixon, 1990; Roberts et al., 1990; Rozario, & Menon, 2010; Roberts & Vernon, 1983). However, a number of other studies employing exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) have yielded different CES-D factor structures, particularly among populations that are
demographically distinct from those where the CES-D was developed. To date, studies have reported two (Cheung & Bagley, 1998; Edman et al., 1999; Kazarian, & Taher, 2010; Rivera-Medina, Caraballo, Rodriguez-Cordero, Bernal, & Davila-Marrero, 2010), three (Guarnaccia, Angel, & Worobey, 1989; Kuo, 1984; Liang, Van Tran, Krause, & Markides, 1989; Yen, Robins, & Lin, 2000; Ying, 1988), and as many as five (Thorson & Powell, 1993; Ying, Lee, Tsai, Yeh, & Huang, 2000) factors. Furthermore, some studies have reported four-factor solutions that significantly differed from Radloff’s original classification (Crockett, Randall, Shen, Russell, & Driscoll, 2005; Foley, Reed, Mutran, & DeVellis, 2002).

Shafer (2006) conducted a relatively recent meta-analysis, which investigated the CES-D factor structure with over 22,000 participants, from 21 CES-D studies. Using EFA, the meta-analysis concluded that the original Radloff four-factor structure was the best solution. However, this study also described reasonable two and three-factor solutions, raising further questions about the universality of the four-factor solution.

More recently, Kim, DeCoster, Huang, and Chiriboga (2011). Race/Ethnicity and the Factor Structure of the Center for (2011) conducted a meta-analysis including 13 EFA studies (N = 19,206) and 16 CFA studies (N = 65,554) that investigated the structure of CES-D across various racial/ethnic groups. They concluded that CFA studies supported the original four-factor structure across all racial/ethnic groups, with the exception of Asian groups. However, results from the EFA meta-analysis did not consistently replicate the original four-factor solution. Therefore, these authors (Kim et al., 2011) recommended that researchers, first, should use EFA and, second, apply CFA when examining the factor structure of the CES-D in racially/ethnically diverse groups. Thus, the current study shall follow their recommendation. Moreover, it is worth noting that the meta-analysis encompassed only seven studies (EFA = 4, CFA = 3) that utilized child and adolescent samples. Consequently, more attention should be paid to the structure of the CES-D within child and adolescent groups, particularly in non-Western cultures.

To date, the results of previous CES-D factor analytic studies within Chinese samples (both adult and adolescent) have been mixed (cf. Cheng et al., 2012; Lee et al., 2008; Li et al., 2010; Lin, Wei, Yi, Xiao, & Yao, 2008; Yen, Robins, & Lin, 2000; Zhang et al., 2011, 2010). Indeed, a study conducted with Mainland Chinese high school students (Lin et al., 2008) reported that the initial four-factor model proposed by Radloff (1977) fitted the data appropriately; however, they failed to compare the four-factor model with other competing models.

Similar results were observed in a mainland general population study (Zhang et al., 2010), which utilized an elderly sample (Zhang et al., 2011), Taiwan adolescents (Cheng et al., 2012), and a sample of children from Hong Kong (Li et al., 2010). However, an alternative study based on Chinese university students, which employed principal components analysis with varimax rotation (Yen, Robins & Lin, 2000), identified three factors. In this solution, items from the original DA, IP, and SC factors merged into two new factors labeled as a somatic factor and an affective factor. This model confirmed the results of previous studies (Liang et al., 1989; Manson, Ackerson, Dick, Baron, & Fleming, 1990) and was recently confirmed in a CFA study utilizing data from Hong Kong adolescents (Lee et al., 2008).

Moreover, several studies have found that two to five factor solutions fit their data best. In a sample of 138 Hong Kong married couples, Cheung and Bagley (1998) reported a two-factor structure that separated the IP items from the remaining items, represented the best fit to the data. Inconsistent with Cheung and Bagley’s two-factor solution, Lam, Stewart, Leung, Lee, and Wong (2004) reported an alternative two-factor solution, with the four positive items loading on a factor separate from the remaining items, in a Hong Kong adolescent sample. However, Ying et al. (2000) reported a five-factor solution in a Chinese American college student sample. Thus, a consensus regarding the underlying dimensionality of the CES-D with the population of China has yet to be reached.

Notably, there may be several limitations to the analytic approach used in previous Chinese sample studies. First, the data collected from the CES-D have often been treated as continuous despite items being based on only four response categories (Cheng et al., 2012; Li et al., 2010; Zhang et al., 2011, 2010). Treating Likert rating scale data, which have four or less response categories, as continuous outcomes violates the assumption of multivariate normality, and thus may distort the resultant factor structure and associated parameters (DiStefano, 2002; Lubke & Muthén,
Second, most of the above-mentioned studies, which are based on Chinese samples, employed CFA to compare several alternative models of the CES-D. This may reflect a missed opportunity to uncover a Chinese-specific structure through the use of EFA (Kim et al., 2011).

Third, although some researchers have tested the measurement invariance of the CES-D between Chinese populations with varying cultural backgrounds (Zhang et al., 2011), no study has examined the measurement invariance of the CES-D across gender in the Mainland China population. Nor has such a study been implemented using an adolescent sample. To overcome those shortfalls, the present study utilizes a representative Mainland Chinese adolescent sample, employing both EFA and CFA to explore the structure of the CES-D.

Indeed, exploring the factor structure and measurement invariance of the CES-D has significant meaning for both clinicians and researchers. From a cross-cultural standpoint, depression is sometimes manifested and expressed differently across various cultural groups and subgroups (e.g., Iwata, Tuner, Lloyd, 2002), so exploring the factor structure of a depression measure (e.g., CES-D) enables us to examine any differences across cultures and groups. Moreover, different factor structures imply variant psychological mechanisms, thus suggesting that diverse therapeutic schedules should be employed for differing people. Specifically, effective therapeutic schedules verified in one culture may be invalid for people from another culture.

The main aim of this study was to examine the factor structure of the CES-D in a large representative Mainland Chinese adolescent sample. Inconsistent with previous studies using Chinese samples (e.g., Cheng et al., 2012; Zhang et al., 2011) and according to the recommendation of Kim et al. (2011), we first employed EFA to explore the structure of the CES-D in sample 1, and then in sample 2 we used CFA to compare the EFA-derived model with other competing models which were shown to be superior in previous studies. Moreover, although no significant gender differences were found in Chinese adolescents based on manifest mean levels (Greenberger, Chen, Tally, & Qi, 2000; Tepper et al., 2008; Yen, Robins, & Lin, 2000), if measurement invariance does not hold across gender, differences in observed depressive symptoms scores may not be directly comparable (Meade, Lautenschlager, & Hecht, 2005). Thus, the second aim was to examine measurement invariance of the best fitting model, identified in the above-mentioned analysis, across gender, and to compare sex differences of depressive symptoms at the latent mean level.

Method

Participants and Procedure
A total of 5,059 participants (48% boys), aged 14 to 20 years, were recruited (mean \( M = 16.55 \), standard deviation \( SD = 1.06 \)). Among study participants, 47.7% were in the United States’ equivalent of Grade 9, with 30.2% in Grade 10, and 22.1% in Grade 11. The sample encompassed 93.4% of individuals who reported their ethnicity as Han. A further 6.6% classified themselves as belonging to an ethnic minority or declared that their ethnicity was unknown. With regards to family composition, participants reported the following: 88.5% declared that their family was a nuclear family, 7.1% reported divorced families, 3.2% reported remarried families, and 2.2% reported single-parent families.

Participants completed the survey in school during a specified class period lasting approximately 45 minutes. Informed consent was given by parents (or legal guardians) prior to the administration of the self-reported questionnaires. Furthermore, children provided assent for their own participation. Ethical approval was given by the Human Subjects Review Committee at Central South University.

Measure

The Chinese version of the CES-D. The CES-D was used to measure the level of depressive symptomatology. The CES-D comprises 16 negative affect and four positive affect items, such as “I felt depressed,” “I felt lonely,” and “I was happy.” Participants were asked about the number of days on which they experienced depressive symptoms during the previous week.
Respondents reported the frequency of occurrence, of each item, during the previous week on a 4-point scale: 0 (rarely; less than 1 day), 1 (some of the time; 1–2 days), 2 (a moderate amount of the time; 3–4 days), or 3 (most or all of the time; 5–7 days). Higher scores on the CES-D indicate more depressive symptoms (Radloff, 1977). The four positive affects items were reversed before conducting our analysis. The Chinese version of this scale has been validated (Cheng et al., 2012; Lee et al., 2008; Zhang et al., 2010) and extensively utilized in Chinese studies (e.g., Yen, Robins, & Lin, 2000; Zhang et al., 2011, 2010).

Data Analysis Strategy

Missing data. The original sample included 5,321 adolescents; however, because 262 failed to respond to all of the CES-D items, they were excluded from the analysis. This left an effective sample size of 5,059.

Analytic steps. Our analyses contained three steps. First, EFA was conducted on a randomly split-half of the whole sample (n = 2526). EFA was used to identify the best fitting factor model of the CES-D, in the present sample. Subsequently, we used a random split-half (n = 2533) of the sample to run CFA. The CFA tested the fit of a number of competing models, including the model that was generated from our EFA. Finally, utilizing the full sample we assessed measurement invariances of the best fitting model, from the CFA, across gender.

Stage 1: EFA. Descriptive statistics and EFA were performed by the SPSS program (IBM, SPSS version 17, 2009). In line with previous studies (e.g., Radloff, 1977; Yen, Robins, & Lin, 2000), principal components analysis with varimax rotation was performed to determine the number of factors. However, given that the oblique rotation approach is also appropriate in this case, we computed the oblique rotation solution. Notably, the factor-loading matrix was similar to that of the varimax rotation solution. Further details and matrices are available from the first author.

Stage 2: CFA. A series of CFAs were specified and estimated using Mplus 5.1 software (Muthén & Muthén, 1998–2007). Given that items have only four response categories, maximum likelihood (ML) estimation was deemed inappropriate in light of simulations studies showing that a minimum of five response categories is a prerequisite to the assumptions of continuity underlying ML estimation (DiStefano, 2002; Lubke & Muthén, 2004). Thus, the robust weighted least squares with mean and variance adjustment (WLSMV) estimator was used in present study (Flora & Curran, 2004).

CFA model specification. Five alternative models of the CES-D, which were shown to be best fitting across previous studies, were chosen for comparison (cf. Table 1). Model 1 tested Radloff’s original four-factor model in which all 20 items loaded on four specific factors: depressed, somatic, interpersonal, and positive. The four-factor model has been shown to be the best fitting model across various Chinese factor analytic studies (Cheng et al., 2012; Lee et al., 2008; Li et al., 2010; Zhang et al., 2011, 2010). Model 2 tested a two-factor model identified in Cheung and Bagley’ study (1998) in which the interpersonal problem items were separated from the remaining items.

Model 3 tested another two-factor model that combined all negative items into an independent factor, while the remaining positive items formed a second factor. Recently, this particular model has been verified as superior in several studies (Edwards, Cheavens, Heiy, & Cukrowicz, 2010; Kazarian, & Taher, 2010; Leykin, Torres, Aguilera, & Muñoz, 2011; Rivera-Medina et al., 2010; Schroepers, Sandersman, van Sonderen, & Ranchor, 2000). Notably, it has been shown to be superior in Hong Kong (Lam et al., 2004) and Filipino-American (Edman et al., 1999) adolescents. Model 4 tested the three-factor model that was identified in EFA (see EFA results section). Model 5 tested an alternative three-factor model encompassing a factor of depressed affect and somatization, a factor of positive affect, and a factor of interpersonal problems (Guarnaccia et al., 1989; Kuo, 1984; Ying, 1988).
Table 1
Item Mapping for Tested Models

<table>
<thead>
<tr>
<th>Item content</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bothered by things</td>
<td>SC</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>2. Appetite was poor</td>
<td>SC</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>3. Can’t shake off the blues</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>4. Just as good as others</td>
<td>PA</td>
<td>DA</td>
<td>PA</td>
<td>PA</td>
<td>PA</td>
</tr>
<tr>
<td>5. Trouble concentrating</td>
<td>SC</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>6. Felt depressed</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>7. Everything was an effort</td>
<td>SC</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>8. Hopeful about the future</td>
<td>PA</td>
<td>DA</td>
<td>PA</td>
<td>PA</td>
<td>PA</td>
</tr>
<tr>
<td>9. Life has been a failure</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>10. Fearful</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>11. Sleep was restless</td>
<td>SC</td>
<td>DA</td>
<td>DA</td>
<td>SC</td>
<td>DA</td>
</tr>
<tr>
<td>12. Happy</td>
<td>PA</td>
<td>DA</td>
<td>PA</td>
<td>PA</td>
<td>PA</td>
</tr>
<tr>
<td>13. Talked less than usual</td>
<td>SC</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
</tr>
<tr>
<td>14. Felt lonely</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
</tr>
<tr>
<td>15. People were unfriendly</td>
<td>IP</td>
<td>IP</td>
<td>DA</td>
<td>DA</td>
<td>IP</td>
</tr>
<tr>
<td>16. Enjoyed life</td>
<td>PA</td>
<td>DA</td>
<td>PA</td>
<td>PA</td>
<td>PA</td>
</tr>
<tr>
<td>17. Had crying spells</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
</tr>
<tr>
<td>18. Felt sad</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
</tr>
<tr>
<td>19. People disliked me</td>
<td>IP</td>
<td>IP</td>
<td>DA</td>
<td>DA</td>
<td>IP</td>
</tr>
<tr>
<td>20. Could not get going</td>
<td>SC</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
<td>DA</td>
</tr>
</tbody>
</table>

Note. DA = depressed affect; IP = interpersonal problem; PA = positive affect; SC = somatic complaints; M1 = Radloff's original four-factor model; M2 = two factor model identified in Cheung and Bagley’s study (1998) in which DA + PA + SC as a factor and IP; M3 = another two-factor model in which all negative items were combined into an independent factor and the remaining four positive items formed a second factor; Model 4 = positive affect and two new factors merged from original depressed affect, interpersonal problem, and somatic complaints; Model 5 = another three-factor model with depressed affect and somatic complaints, positive affect, and interpersonal problem.

Model evaluation in CFA. Following generally accepted practice, we evaluated the fit of each model by examining multiple fit indices (Kline, 2010). We used the chi-square, root mean square error of approximation (RMSEA), Tucker-Lewis index (TLI), and comparative fit index (CFI). Conventional guidelines suggest that RMSEA values ≤ .08 indicate acceptable model fit and ≤ .05 indicate good model fit, and CFI, TLI ≥ .90 indicate adequate model fit (Kline, 2010).

Regular chi-square difference tests were not conducted here for the comparison of non-nested competing models. Rather, we employed the Bayesian information criterion (BIC) as it was most suitable in this case. A 10-point BIC difference represents a 150:1 likelihood (p < .05) that the model with the lower BIC value fits best; a difference in the 6–10-point range indicates “strong” support, and > 10 indicates “very strong” support (Raftery, 1995). However, because the BIC value is not given when utilizing the WLMSV estimator in Mplus, we computed the BIC value by estimating the models using ML estimator.

Stage 3: Measurement invariance. After identification of the best fitting model of the CES-D, we conducted measurement invariance tests across gender. Measurement invariance tests were performed using the sequential strategy described by Meredith and Teresi (2006). First, Model A tested configural invariance by allowing all parameters to vary. Subsequent models tested sequentially more conservative restrictions of Model A by constraining particular parameter estimates to be equal across groups. Subsequent models were tested against the prior step’s model (except when noted otherwise). Model B constrained factor loadings across groups (testing weak invariance). Model C additionally constrained observed variable thresholds (testing strong factorial invariance). Model D additionally constrained item error variances (testing strict factorial invariance). Model E, tested structural invariance by additionally constraining
factor variances and covariances (but not item error variances), while being tested against Model C. And, finally, Model F additionally constrained factor means (but not item error variances), while being tested against Model E. Thus, we tested differences not only in symptom structure between the boys and girls but also in the level of symptom severity.

**Model comparison in measurement invariance.** Because previous research has reported that comparing nested models solely on the bases of fit indices might lead to inaccurate conclusions (Fan & Sivo, 2009), nested models in the current study were compared using the *difftest* function available in Mplus (Muthén & Muthén, 1998–2007). However, given that tests of the change in CFI are reported as being superior to chi-square difference tests of nested models, because they are not affected by the sample size (Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008), the current study compared nested models in consideration of both the chi-square difference test and CFI values. According to Cheung and Rensvold (1999, 2002), a CFI difference of < .01 indicates that the invariance hypothesis should not be rejected as mean differences exist when CFI differences are .01 to .02, and definite differences exist when CFI differences are > .02 (Cheung & Rensvold, 1999, 2002).

**Results**

**Stage 1: Exploratory Factor Analysis**

According to the suggestion offered by Kim et al. (2011), EFA was conducted first on a random split-half of a whole dataset to choose the best factor model representing the structure of CES-D in the present population. This analysis yielded three factors/components with eigenvalues greater than 1.0; moreover, the scree plot test also suggested the retention of three factors. The three factors cumulatively accounted for 48.58% of the total variance. As can be seen in Table 2, the first factor explained 19.89% of the total variance and comprises nine items that belong to the original DA (four items: 3, 6, 9, 10) and SC (six items: 1, 2, 5, 7, 11, 13) factors,
Table 3

<table>
<thead>
<tr>
<th>Model</th>
<th>WLSMV$\chi^2$</th>
<th>df</th>
<th>p</th>
<th>TLI</th>
<th>CFI</th>
<th>BIC</th>
<th>RMSEA (90% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>2183.48</td>
<td>164</td>
<td>&lt;0.01</td>
<td>.949</td>
<td>.956</td>
<td>99875.173</td>
<td>.070 [.068–.071]</td>
</tr>
<tr>
<td>M2</td>
<td>5350.953</td>
<td>169</td>
<td>&lt;0.01</td>
<td>.873</td>
<td>.887</td>
<td>100941.619</td>
<td>.110 [.108–.113]</td>
</tr>
<tr>
<td>M3</td>
<td>2817.744</td>
<td>169</td>
<td>&lt;0.01</td>
<td>.935</td>
<td>.942</td>
<td>100387.025</td>
<td>.079 [.076–.081]</td>
</tr>
<tr>
<td>M4</td>
<td>2261.602</td>
<td>167</td>
<td>&lt;0.01</td>
<td>.948</td>
<td>.954</td>
<td>99987.883</td>
<td>.070 [.068–.073]</td>
</tr>
<tr>
<td>M4 modified</td>
<td>1866.193</td>
<td>165</td>
<td>&lt;0.01</td>
<td>.957</td>
<td>.963</td>
<td>99797.885</td>
<td>.064 [.061–.066]</td>
</tr>
<tr>
<td>M5</td>
<td>2179.399</td>
<td>167</td>
<td>&lt;0.01</td>
<td>.950</td>
<td>.956</td>
<td>99862.198</td>
<td>.069 [.066–.072]</td>
</tr>
<tr>
<td>M6</td>
<td>1354.121</td>
<td>132</td>
<td>&lt;0.01</td>
<td>.961</td>
<td>.966</td>
<td>91410.876</td>
<td>.060 [.058–.063]</td>
</tr>
</tbody>
</table>

Note. WLSMV = weighted least squares with mean and variance adjustment; df = degree of freedom; TLI = Tucker-Lewis index; CFI = comparative fit index; BIC = Bayesian information criterion; RMSEA = root mean square error of approximation; CI = confidence interval; M1 = Radloff’s original four-factor model; M2 = two-factor model identified in Cheung and Bagley’s study (1998) in which DA + PA + SC as a factor and IP; M3 = another two-factor model in which all negative items were combined into a independent factor and remaining four positive items formed a second factor; M4 = positive affect and two new factors merged from original depressed affect, interpersonal problem, and somatic complaints; M4-modified = based on M4 and allow items 9 and 18 cross-loading; M5 = another three-factor model with depressed affect and somatic complaints, positive affect, and interpersonal problem; M6 = based on M4 but deleted items 9 and 18.

as well as two cross-loading items 9 and 18. The second factor explained 16.95% of the total variance and comprises seven items that belong to the original DA (three items: 14, 17, 18), IP (two items: 15, 19), and SC (two items: 13, 20) factors. The last factor explained 11.74% of the total variance and comprises four positive affect items. These results are consistent with findings from other studies (Liang et al., 1989; Yen, Robins, & Lin, 2000; Manson et al., 1990).

The three-factor structure found in the EFA in the current study differs from Radloff’s (1977) factors. In particular, the IP factor dropped out in this study. Additionally, four of seven DA items loaded onto the SC factor and two of seven DA items loaded onto a new DA factor. It was worth noting that the two items, item 9 and item 18, cross-loaded onto more than one factor. Item 9 primarily loaded (loading size = .473) onto SC factor and peripherally loaded onto two other factors (factor loadings were .378 and .376, respectively). Item 18 primarily loaded (factor loading = .559) onto the SC factor and peripherally loaded onto the DA factor (factor loading = .503). Notably, a previous study found similar results (i.e., Yen, Robins, & Lin, 2000).

Stage 2: Confirmatory Factor Analysis

Table 3 summarizes the fit indices of these competing models using the WLSMV estimation method for the polychoric correlation matrix of the CES-D items in the second split-half of the full sample. As can be seen in Table 3, all of the examined models fit the data well (CFIs > .90, TLI s > .90, RMSEAs < .08), with the exception of Model 2 (WLSMV$\chi^2$ = 5350.953, CFI = .887 < .90, TLI = .873 < .90, RMSEA = .11 > .08, BIC = 100941.619), in which IP items were separated from the remaining items. Overall, Model 5 provided the best fit to the data among the five alternative models (WLSMV$\chi^2$ = 2179.399, degree of freedom [df] = 167, CFI = .956, TLI = .950, RMSEA = .069, BIC = 99862.198), in which the DA factor and the SC factor were collapsed into a single factor, in addition to PA and IP factors. The second best fitting model was Radloff’s original four-factor model (WLSMV$\chi^2$ = 2183.48, df = 164, CFI = .956, TLI = .949, RMSEA = .070, BIC = 99875.173). The difference in the resultant BIC value between Model 5 and Radloff’s model was 12.975 (>10), thus suggesting that Model 5 fit the data significantly better than Radloff’s model.

The model identified in the above EFA, Model 4, also showed acceptable fit to the data. However, because Model 4 included two cross-loading items (i.e., item 9 and item 18), we hypothesized this might result in a worse fit and so we checked the modification indices (MI) of
Factor Structure and Invariance of the CES-D

Table 4

<table>
<thead>
<tr>
<th>Model</th>
<th>WLSMV $\chi^2$</th>
<th>df</th>
<th>TLI</th>
<th>CFI</th>
<th>RMSEA (90% CI)</th>
<th>$\Delta\chi^2$</th>
<th>$\Delta df$</th>
<th>$\Delta TLI$</th>
<th>$\Delta CFI$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-Configural invariance</td>
<td>2740.835**</td>
<td>264.958</td>
<td>.963</td>
<td>.061</td>
<td>[.059–.063]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-Weak invariance</td>
<td>2849.063**</td>
<td>279.958</td>
<td>.962</td>
<td>.060</td>
<td>[.058–.062]</td>
<td>169.341**</td>
<td>15</td>
<td>.000</td>
<td>−.001</td>
</tr>
<tr>
<td>C-Scalar invariance</td>
<td>3233.528**</td>
<td>312.957</td>
<td>.958</td>
<td>.061</td>
<td>[.059–.063]</td>
<td></td>
<td></td>
<td>−.001</td>
<td>−.004</td>
</tr>
<tr>
<td>D-Error Variance invariance</td>
<td>2942.618**</td>
<td>330.961</td>
<td>.964</td>
<td>.056</td>
<td>[.054–.058]</td>
<td>466.026**</td>
<td>33</td>
<td>.000</td>
<td>−.001</td>
</tr>
<tr>
<td>E-Variance-covariances</td>
<td>2520.412**</td>
<td>318.967</td>
<td>.969</td>
<td>.052</td>
<td>[.050–.054]</td>
<td>14.574*</td>
<td>6</td>
<td>.006</td>
<td>+.005</td>
</tr>
<tr>
<td>F-Latent mean invariance</td>
<td>2512.446**</td>
<td>321.968</td>
<td>.969</td>
<td>.052</td>
<td>[.050–.054]</td>
<td></td>
<td></td>
<td>.001</td>
<td>.000</td>
</tr>
</tbody>
</table>

Note. WLSMV = weighted least squares with mean and variance adjustment; df = degree of freedom; TLI = Tucker-Lewis index; CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; $\Delta \chi^2$ = change in $\chi^2$ relative to the preceding model; $\Delta df$ = change in degree of freedom relative to the preceding model; $\Delta CFI$ = change in comparative fit index relative to the preceding model; $\Delta TLI$ = change in Tucker-Lewis index relative to the preceding model. Chi-square difference test with WLSMV estimation is different from the conventional chi-square difference test; more information on this can be found at: http://www.statmodel.com/chidiff.shtml.

M4. Indeed, the MIs indicted that if item 9 loaded onto the PA factor and item 18 loaded onto the SC factor, then the chi-square value would reduce 267.102 and 126.477, respectively. Thus, based on the MIs we allowed item 9 to load onto the PA factor and item 18 to load onto the SC factor. We then estimated this model using CFA. The fit indices of the modified model (M4-modified) indicated excellent fit (CFI = .963, TLI = .957, RMSEA = .064, BIC = 99797.885), albeit with a significant chi-square value, WLSMV $\chi^2$ (165, N = 2530) = 1866.193, $p < .001$. All factor loadings were significant at $p < .001$.

We also employed an alternative analytic procedure to allow for a simple structure and lucid explainable results; thus, we created a new version of model 4: M6. This model was essentially identical to the above except we excluded items 9 and 18. However, to ensure we were following correct procedures, we first calculated the Pearson product-moment correlation between original factor scores and new factor scores without the two items. The coefficients were .990 for SC and .986 for DA, respectively, which suggested that the two removed items did not attenuate the correlations between factors. We once again estimated this new model, M6, using CFA. The fit indices of M6 exhibited an excellent fit (WLSMV $\chi^2$ = 1354.121, $df$ = 132, CFI = .966, TLI = .961, RMSEA = .060, BIC = 91410.876). Solely from the point of model fitting, M6 was the best fitting model compared with other alternative models. However, the difference in fit between M6 and the modified M4 was minimal, but the differences between Model 4 and Model 6, in terms of the resultant BIC values, was 8387.007 (>10), thus suggesting that Model 6 fit the data significantly better than Model 4.

Stage 3: Factorial Invariance Across Gender

The results (see Table 4) from the measurement invariance revealed that all of the five steps of invariance testing resulted in significant $\chi^2$ (all $p < .01$), excellent (CFIs > .95, TLIIs > .95), and equivalent fit indices ($\Delta$CFIs < .01, $\Delta$TLIs < .01), with significant $\Delta \chi^2$ (all $p < .01$ expect for Model E, $p < .05$). All goodness-of-fit indices, except for $\chi^2$ and $\Delta \chi^2$, which are sensitive to large sample size, indicated that all models assuming different degrees of invariance were acceptable.
The latent means for the three factors in the second group (i.e., girls; the latent mean for first group was fixed to 0 in order for the model to be identified) were .224 \( (p < .01) \), \( -.109 (p < .01) \), and \( .036 (p > .05) \), respectively. Although the equivalent fit indices (\( \Delta \text{CFIs} < .01, \Delta \text{TLIs} < .01 \)) revealed that the latent mean differences were negligible, we also computed the standardized effect size (Hancock, 2001; Thompson & Green, 2006). The results showed that standardized effect size for the three factors were \( .269 \), \( .296 \), and \( .056 \), respectively. According to Cohen’s (1988), those effect sizes implied that the latent mean differences were small.

In summary, these results suggest that the CES-D items have the same meanings across gender, manifest mean levels (i.e., based on summed/averaged scores), and latent mean comparisons, suggesting that comparisons across sex are meaningful.

Discussion

The primary purpose of this study was to test the factor structure of the CES-D using EFA and CFA among Mainland China adolescents. EFA results revealed that the 20 items of the CES-D can be interpreted in terms of three symptom dimensions, including SC, DA, and PA. This result is partly consistent with EFA findings from other studies (Yen, Robins, & Lin, 2000; Manson et al., 1990), which suggest original three dimensions (SC, IP, and DA) collapsing into two factors: SC and DA. However, this finding is inconsistent with previous researches who have employed CFA to compare alternative structures of depressive symptoms among Chinese populations (Cheng et al., 2012; Zhang et al., 2011).

To validate the structure of the CES-D confirmed in the first split-half of the sample using EFA, we specified and estimated several alternative models that have been previously reported as best fitting across a number of studies. Models were tested using CFA in the second split-half of the sample. The EFA model that we found was similar to another three-factor model (Lee et al., 2008; Liang et al., 1989; Manson et al., 1990; Yen, Robins, & Lin, 2000) in which DA and SC merged into a single factor. However, this model included two cross-factor loadings, and, thus, when we tested it using CFA, it fit adequately but not excellently. Therefore, we removed the items that had cross-factor loading and respecified the model.

This newer modified model was found to fit the data best compared with alternative models, and compared with the model found in EFA, which had cross-factor loadings. However, to allow for simple structure and lucid explainable result, we planned to delete those cross-loading items (i.e., items 9 and 18). Thus, a new model labeled M6 was tested, which resulted in best fit. To guarantee this decision, we calculated the Pearson product-moment correlation between factor scores with the two items and scores without them and another EFA. Consequently, a three-factor model best represented the structure of the 18 items of CES-D among Chinese adolescents.

Furthermore, the findings of the EFA and CFA together suggest that there is a specific structure of depressive symptoms in Mainland China adolescents that differs from the structure found by recent alternative studies of Chinese adolescents (Cheng et al., 2012; Lin et al., 2008; Zhang et al., 2010). The structure also differs from that found in other adolescent populations from varying cultural backgrounds (Edman et al., 1999; Radloff, 1991; Roberts, & Sobhan, 1992). As mentioned above, prior studies (Cheng et al., 2012; Li et al., 2010; Lin et al., 2008; Zhang et al., 2011, 2010) that were conducted using Chinese populations have failed to explore the structure of depressive symptoms prior to conducting the CFA (Kim et al., 2011).

Additionally, CFA’s have employed inappropriate estimators when conducting the CFA (e.g., ML). Notably, a recent meta-analysis (Kim et al., 2011) proposed that when testing the structure of depressive symptoms in new cultures, researchers should first implement EFA and then move to CFA. Indeed, they suggested that conducting CFA as a first point of call may be misleading. Moreover, although the original four-factor model of the CES-D provided reasonable fit to the data using CFA in the current study, it was not found in the EFA. Consequently, our results together with Kim et al.’s (2011) suggestion support the implementation of EFA prior to CFA in the current study.

Our results also showed that the three-factor model that merged DA and SC factors into one single factor, in addition to PA and IP factors, fit the data best (that is prior to the testing of
Factor Structure and Invariance of the CES-D 975

modified models). The fact that DA and SC factors intertwine implies that Mainland China adolescents are prone to mingle somatization with the expression of depressive mood. In other words, Chinese adolescents tend to express somatic symptoms more so than psychological symptoms of depression as more DA items were treated as SC items. This result is in agreement with traditional cultural views that Chinese people are encouraged not to express their mood condition, i.e., psychological symptoms; thus, somatization is an alternative way to express emotional disturbance (Kleinman, 1982; Cheung, 1995).

Furthermore, given that mental illnesses are stigmatized in Chinese societies (Chan & Parker, 2004; Chung & Wong, 2004; Parker, Gladstone, & Chee, 2001; Ryder, Bean, & Dion, 2000), this may further discourage the expression of psychological symptoms. Consequently, somatic symptoms are emphasized in Chinese people (Cheung, 1995; Parker et al., 2001; Ryder et al., 2008). Interestingly, despite rapid modernization in Mainland China across the last decade, the way in which mental illness is expressed, particularly depression, has not changed (Kleinman, 1982).

The divergence among studies with respect to the structure of the CES-D in Chinese populations may be partially attributed to culture influence. Ying et al.’s (2000) study revealed that SC and DA were separate factors in a sample of Chinese-American college student ($n = 353$), 34.56% of which were American-born, and the rest were immigrants. However, in studies of other Chinese populations, i.e., those who were born and raised in China (Yen et al., 2000) and those who born in China and subsequently immigrated to the United States (Ying, 1988), results have concluded that the SC and DA items tend to mix. To the best of our knowledge, until now, no one study has directly addressed this question. Thus, future cross-cultural studies, considering explicit cultural variables (e.g., individualism vs. collectivism; cultural identity), are needed to determine the cultural boundaries of the model.

The second aim of the current study was to examine measurement invariance across gender and to compare the sex difference of depressive symptoms on latent mean level. Previous studies have found that sex has influence on the factor structure of the CES-D (Clark, Aneshensel, Frerichs, & Morgan, 1981; Guarnaccia et al., 1989). In the current Chinese adolescent sample, however, we failed to replicate this result. All models assuming different degrees of invariance were acceptable, suggesting that the CES-D’s factors have the same meaning across gender, as well as manifest mean levels (i.e., based on summed/averaged scores), which suggests that comparisons across sex are meaningful. Our findings were consistent with Cheng et al.’s (2012) investigation, in which no measurement noninvariance was found in a sample of Taiwan adolescents.

As mentioned above, no significant gender differences were reported in Chinese adolescents with regard to manifest mean levels (Greenberger et al., 2000; Tepper et al., 2008; Yen, Robins, & Lin, 2000), and, thus, our findings supply additional empirical support for their conclusion. It is worth noting that previous studies have compared the depressive symptoms of Chinese people with people from various cultural backgrounds (e.g., Greenberger et al., 2000; Parker et al., 2001; Ryder et al., 2008; Yen, Robins, & Lin, 2000). However, as related to gender, if the measurement invariance does not hold across groups, differences in observed scores might not be directly comparable. Thus, it is crucial that future studies take measurement invariance into consider when conducting cross-cultural research.

Our results also have crucial meaning for clinicians. First, Mainland China adolescents are prone to mingle somatization with the expression of depressive mood, which remind the clinicians to be cautious with adolescent outpatients and to be wary that they do not obscure and/or miss potential depressed adolescents. Second, specific therapeutic schedules should be developed and employed for this group. Finally, it is significant that future depression measurement and diagnostic criteria should take this character into consider.

The current study is not without limitation. For example, we employed a restricted sample of urban adolescents; thus, the results may not fully generalize to all Chinese adolescents, in particular those from rural parts of China. Furthermore, our sample was not a clinical one and therefore may not generalize well to individuals who receive clinical diagnoses of depression. More positively, however, our study employed a large-scale, representative urban sample. In
addition, we used a more appropriate parameter estimate approach (Flora & Curran, 2004). These aspects will have enhanced the research relevance of our results.

Conclusion

The results of the present investigation indicated that the original four-factor model of the CES-D was unsuitable for Mainland Chinese adolescents. Indeed, results suggested that three factors better represented the underlying structure of CES-D. Notably, the resultant best fitting, three-factor structure intertwined DA and SC items. In addition, results suggested the best fitting model was one that deleted two of the CES-D items. Furthermore, the three-factor structure was shown to be invariant across subgroups of boys and girls. Therefore, the current study provides further evidence that boys and girls have a uniform structure of depression symptoms; in other words, no significant gender differences were found.

References


Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. Medical Care, 44(11, suppl 3), S69–S77. doi:10.1097/01.mlr.0000245438.73837.89


