Running Head: FACTOR STRUCTURE AND INVARIANCE OF THE CES-D

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

Mengcheng Wang

Central South University, Changsha, China

Cherie Armour

University of Southern Denmark, Funnen, Denmark

Yan Wu

Guangdong University of Foreign Studies, Guangzhou, China

Fen Ren

Institute of Psychology, Chinese Academy of Sciences, Beijing, China

Xiongzhao Zhu, and Shuqiao Yao

Central South University, Changsha, China

Author Note

Mengcheng Wang, Xiongzhao Zhu, and Shuqiao Yao, Medical Psychological Institute, Second Xiangya Hospital, Central South University, Changsha, China; Cherie Armour, National Centre for Psychotraumatology, University of Southern Denmark, Odense Campus, Funnen, Denmark. Yan Wu, Department of Applied Psychology, Guangdong University of Foreign Studies, Guangzhou, China. Fen Ren, Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China.

Correspondence concerning this article should be addressed to Shuqiao Yao, Medical Psychological Institute, Second Xiangya Hospital, Central South University, #139 Ren-Min Zhong Road, Changsha, 410011, China. Phone: 86-731-85292130, FAX: 86-731-85361328, E-mail: Shuqiaoyao@163.com

Abstract

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 and confirmatory factor analysis were simultaneously conducted to determine the structure of the CES-D in a large scale, representative adolescent samples recruited from mainland China. Multiple-group confirmatory factor analysis (N = 5059, 48% boys, M = 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 three 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.

Key Words: Depressive symptom; CES–D; factor structure; factorial invariance; Chinese adolescent

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

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, Fokkema, Cuijpers, Smits, & Beekman, 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, 7 items), Somatic Complaints (SC, 7 items), Interpersonal Problems (IP, 2 items), and Positive Affect (PA, 4 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, Danko, Andrade, McArdle, Foster, & Glipa, 1999; Kazarian, & Taher, 2010; Rivera-Medina, Caraballo, Rodrıguez-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 and colleagues (2011) conducted a meta-analysis including 13 exploratory factor analytic studies (N = 19, 206) and 16 confirmatory factor analytic 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 should first use EFA and secondly 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 only encompassed seven studies (EFA = 4, CFA = 3) which 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, Chung, & Ho, 2010; Lin, Wei, Yi, Xiao, & Yao, 2008; Yen, Robins, & Lin, 2000; Zhang, Fokkema, et al., 2011; Zhang, Wu, Fang, Li, Han, & Chen, 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, Wu, et al., 2010), a study utilizing an elderly sample (Zhang, Fokkema, et al., 2011), Taiwan adolescents (Cheng et al., 2012), and a sample of children from Hong Kong (Li, Chung, & Ho, 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 et al., 1990) and was recently confirmed in a confirmatory factor analytic 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 et al. (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 has often been treated as continuous despite items being based on only four response categories (Cheng et al., 2012;

Li, Chung, & Ho, 2010; Zhang, Fokkema, et al., 2011; Zhang, Wu, et al, 2010). Treating Likert rating scale data, which has four, or less than four, 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, 2004). 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, Fokkema, et al., 2011), no study has examined the measurement invariance of the CES-D across gender in the mainland China population. Nor has such been implemented using an adolescent sample. In order 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 et al., 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, Fokkema, 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 one, and then in sample two 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, Liu, Guo, Zhai, Liu, & Li, 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 above analysis, across gender, and to compare sex differences of depressive symptoms at the latent mean level.

Method

Participants and Procedure

A total of 5059 participants (48% boys) were recruited whose ages ranged from 14 to 20 (M = 16.55, 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 consists of 16 negative affect and 4 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, Wu, et al., 2010) and extensively utilized in Chinese studies (e.g., Yen, Robins, & Lin, 2000; Zhang, Fokkema, et al., 2011; Zhang, Wu, et al., 2010).

Data analysis strategy

Missing Data

The original sample included 5321 adolescents, however given that 262 failed to respond to all of the CES-D items they were excluded from the analysis. This left an effective sample size of 5059.

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 which 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: Exploratory Factor Analysis

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: Confirmatory Factor Analysis

A series of confirmatory factor analyses 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 are 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, Chung,

& Ho, 2010; Zhang, Fokkema, et al., 2011; Zhang, Wu, et al., 2010). Model 2 tested a two factor model identified in Cheung and Bagley' study (1998) in which the interpersonal problems items were separated from the remaining items. Model 3 tested another two factor model that combined all negative items into an independent factor whilst 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; Schroevers et al., 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 3 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).

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, given that 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 residual error variances (testing strict factorial invariance). Model E, tested structural invariance by additionally constraining factor variances and covariances (but not error variances), tested against Model C; and finally, Model F, additionally constrained factor means (but not residual variances), while being tested against Model E. Thus, we not only tested differences in symptom structure between the boys and girls, but also differences in the level of symptom severity.

Model comparison in Measurement Invariance

Given that previous research has reported that comparing nested models solely on the bases of fit indices may 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 firstly on a random split half of whole data-set to choose the best factor model representing the structure of CES-D in present population. This analysis yielded 3 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 1, the first factor explained 19.89% of the total variance and consists of nine items that belong to the original DA (4 items: 3, 6, 9, 10) and SC (6 items: 1, 2, 5, 7, 11, 13) factors, as well as two cross-loading items 9 and 18. The second factor explained 16.95% of the total variance and consists of 7 items that belong to the original DA (3 items: 14, 17, 18), IP (2 items: 15, 19), and SC (2 items: 13, 20) factors. The last factor explained 11.74% of the total variance and consists of 4 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 3 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 nine 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, TLIs > .90, RMSEAs < .08) with the exception of Model 2 (WLSMVχ^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 (WLSMVχ^2=2179.399, *df* = 167, CFI = .956, TLI = .950, RMSEA = .069, BIC= 99862.198), in which the DA factor and the SC factor were collapsed into single factor, in addition to PA and IP factors. The second best fitting model was Radloff's original four-factor model (WLSMVχ^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, given that Model 4 included two cross-loading items (i.e., item 9 and item 18) we hypothesized this may result in worse fit therefore we checked the modification indices (MI) of M4. Indeed, the MI's indicted that if item 9 loaded onto the PA factor and item 18 loaded onto the SC factor, the chi square value would reduce 267.102 and 126.477, respectively. Thus, based on the MI's 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, WLSMVχ^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 however we excluded items nine and 18. However, to ensure we were following correct procedures we firstly 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, suggesting that the two removed items did not attenuate the correlations between factors. We once again estimated this new model, model M6 using CFA. The fit indices of M6 exhibited an excellent fit (WLSMVχ^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 to 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*s < .01), excellent (CFIs > .95, TLIs > .95) and equivalent fit indices (ΔCFIs < .01, ΔTLIs < .01), with significant $\Delta\chi 2$ (all *p*s < .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. 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 twenty 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, Yen, et al., 2011; Zhang, Fokkema, 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 which 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 which 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 which had cross-factor loading and re-specified the model. This newer modified model was found to fit the data best, compared to alternative models, and compared to 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., item 9, 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 factors model best represented the structure of the 18 items of CES-D among Chinese adolescents.

Furthermore, the findings of the EFA and CFA together suggests that there is a specific structure of depressive symptoms in mainland China adolescents which differs from the structure found by recent alternative studies of Chinese adolescents (Cheng et al., 2012; Lin et al., 2008; Zhang, Wu, 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, Chung, & Ho, 2010; Lin et al., 2008; Zhang, Fokkema, et al., 2011; Zhang, Wu, et al., 2010) which have been 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 which 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 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, Gladstone, & Chee, 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.' (2000) study revealed that SC and DA were separate factors in a sample of Chinese-American college student (n = 353), of which 34.56% 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 in order 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, Angel, & Worobey, 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). Thus, suggesting that comparisons across sex are meaningful. Our findings were consistent with Cheng et al.s' (2012) investigation, in which no measurement non-invariance were 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, Gladstone, & Chee, 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 may 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. Firstly, 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. Secondly, 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 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 4-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 which 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.

Acknowledgments:

This study is supported by grants from the National Key Technologies R&D Program in the 11th 5-year plan of China (grant no. 2009BAI77B02) and partly by a grant (in part) from Guangdong Province, philosophy and social sciences "Second Five" planning project (grant no.GD11CXL03) and partly by the 12^{th} 5-year plan of educational sciences of Hunan province, (grant no. XJK011BXL002).

Reference

Chan, B., & Parker, G. (2004). Some recommendations to address depression in Chinese people in Australasia. *Australian and New Zealand Journal of Psychiatry, 38*, 141–147.

Cheung, C., & Bagley, C. (1998). Validating an American scale in Hong-Kong: The Center for Epidemiological Studies Depression scale (CES-D). *The Journal of Psychology, 132*, 169–186. doi:10.1080/00223989809599157

Cheng, C. P., Yen, C. F., Ko, C. H., & Yen, J. Y.(2012). Factor structure of the center for epidemiologic studies depression scale in Taiwanese adolescents. *Comprehensive Psychiatry, 53*, 299-307. doi:10.1016/j.comppsych.2011.04.056

Cheung, F. M. (1995). Facts and myths about somatization among the Chinese. In T.-Y. Lin, W. S. Tseng, & E. K. Yeh (Eds.), Chinese societies and mental health (pp. 156–180). Hong Kong: Oxford University Press.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1–27. doi:10.1177/014920639902500101

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. doi:10.1207/S15 328007SEM0902_5

Chung, K. F., &Wong, M. C. (2004). Experience of stigma among Chinese mental health patients in Hong Kong. *Psychiatric Bulletin, 28*, 451–454. doi:10.1192/pb.28.12.451

Clark, V. A., Aneshensel, C. S., Frerichs, R. R., Morgan, T. M. (1981). Analysis of effects of sex and age in response to items on the CES-D scale. *Psychiatry Research, 124*, 171-81. doi:10.1016/0165-1781(81)90047-0

Crockett, L., Randall, B., Shen, Y., Russell, S., & Driscoll, A. (2005). Measurement equivalence of the Center for Epidemiological Studies Depression Scale for Latino and

Anglo adolescents: A national study. *Journal of Consulting and Clinical Psychology, 73*, 47–58. doi:10.1037/0022-006X.73.1.47

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9*, 327-346. doi:10.1207/S15328007SEM0903_2

Edman, J. L., Danko, G. P., Andrade, N., McArdle, J. J., Foster, J., & Glipa, J.(1999). Factor structure of the CES-D (Center for Epidemiologic Studies Depression Scale) among Filipino-American adolescents. *Social Psychiatry and Psychiatric Epidemiology, 34*, 211–215. doi:10.1007/s001270050135

Edwards, M. C., Cheavens, J. S. Heiy, J. E., & Cukrowicz, K. C. (2010). A Reexamination of the Factor Structure of the Center for Epidemiologic Studies Depression Scale: Is a One-Factor Model Plausible? *Psychological Assessment, 22*, 711–715. doi: 10.1037/a0019917

Fan, X., & Sivo, S. A. (2009). Using ΔGoodness-of-Fit Indexes in Assessing Mean Structure Invariance. *Structural Equation Modeling, 16*, 54–69. doi: 10.1080/10705510802561311

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491. doi:10.1037/1082-989X.9.4.466

Foley, K. L., Reed, P. S., Mutran, E. J., & DeVellis, R. F. (2002).Measurement adequacy of the CES-D among a sample of older African-Americans. *Psychiatry Research, 109*, 61–69. doi:10.1016/S0165-1781(01)00360-2

Golding, J., & Aneshensel, C. (1989). Factor structure of the Center for Epidemiologic Studies Depression Scale among Mexican Americans and non-Hispanic Whites. *Psychological Assessment, 1*, 163–168. doi:10.1037//1040-3590.1.3.163

Graham, J.W. (2009).Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology, 60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530

Greenberger, E., Chen, C., Tally, S.R., & Qi, D. (2000). Family, peer, and individual correlates of depressive symptomatology among U.S. and Chinese adolescents. *Journal of Consulting and Clinical Psychology, 68*, 209–219. doi:10.1037//0022-006X.68.2.209

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 Science and Medicine, 29*, 85-94. doi:10.1037//0022-006X.68.2.209

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. *Psychological Assessment, 2*, 64–72.

Iwata, N., Tuner, R.J., Lloyd, D.A. (2002). Race/ethnicity and depressive symptoms in community-dwelling young adults: a differential item functioning analysis. *Psychiatry Research, 110*, 281–289. doi:10.1016/S0165-1781(02)00102-6

Kazarian, S. S.. & Taher, D.(2010). Validation of the Arabic Center for Epidemiological Studies Depression (CES-D) Scale in a Lebanese Community Sample. *European Journal of Psychological Assessment, 26*, 68–73. doi: 10.1027/1015-5759/a000010

Kleinman, A. (1982). Neurasthenia and depression: A study of somatization and culture in China. *Culture, Medicine and Psychiatry, 6*, 117–190. doi:10.1007/BF00051427

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, New York: Guilford Press.

Kim, G., DeCoster, J., Huang, C., & Chiriboga, D. A.(2011). Race/Ethnicity and the Factor Structure of the Center for Epidemiologic Studies Depression Scale: A Meta-Analysis. *Cultural Diversity and Ethnic Minority Psychology, 17*, 381–396. doi: 10.1037/a0025434

Kuo, W. H. (1984). Prevalence of depression among Asian-Americans.*Journal of Nervous* *and Mental Disease, 172*, 449–457. doi:10.1097/00005053-198408000-00002

Lam, T. H., Stewart, S. M., Leung, G. M., Lee, P. W. H., & Wong, J. P. S. (2004). Depressive symptoms among Hong Kong adolescents: Relation to atypical sexual feelings and behaviors, gender dissatisfaction, pubertal timing, and family and peer relationships. *Archives of Sexual Behavior, 33,* 487–496. doi:10.1023/B:ASEB.0000037429.79394.c6

Lee, S. W., Stewart, S. M., Byrne, B. M., Wong, J., Ho, S. Y., Lee, P., & Lam, T. H.(2008). Factor Structure of the Center for Epidemiological Studies Depression Scale in Hong Kong Adolescents. *Journal of personality assessment, 9*0, 175-184. doi:10.1080/00223890701845385

Leykin, Y., Torres, L. D., Aguilera, A., & Muñoz, R. F. (2011). Factor structure of the CES-D in a sample of Spanish- and English-speaking smokers on the Internet. *Psychiatry Research, 185*, 269–274. doi:10.1016/j.psychres.2010.04.056

Liang, J., Van Tran, T., Krause, N., & Markides, K. (1989). Generational differences in the structures of the CES-D in Mexican Americans. J*ournal of Gerontology, 44,* 110–120.

Li, H. C. W., Chung, O. K. J., & Ho, K. Y.(2010).Center for Epidemiologic Studies Depression Scale for Children: psychometric testing of the Chinese version. *Journal of Advanced Nursing, 66*, 2583-2591.

Lin, Y., Wei, Y., Yi, J., Xiao, J., & Yao, S.(2008). Factorial Structure of the CES-D Scale Among Chinese High School Students. *Chinese Journal of Clinical Psychology, 16*, 265-267. (In Chinese).

Lubke, G. H., & Muthen, B. O. (2004). Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling, 11*, 514-534. doi:10.1207/s15328007sem 1104_2

Manson, S. M., Ackerson, L. M., Dick, R. W., Baron, A. E., & Fleming, C. M. (1990).

Depressive symptoms among American Indian adolescents: Psychometric characteristics of the Center for Epidemiologic Studies Depression Scale (CES–D). *Psychological Assessment, 2*, 231–237.

Meade, A. M., Lautenschlager, G. J., & Hecht, J. E. (2005). Establishing measurement equivalence and invariance in longitudinal data with item response theory. *International Journal of Testing, 5*, 279-300. doi:10.1207/s15327574ijt0503_6

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*, 568-592. doi: 10.1037/0021-9010.93.3.568

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

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, California: Muthén & Muthén.

Parker, G., Cheah, Y. C., & Roy, K. (2001). Do the Chinese somatize depression? A cross-cultural study. *Social Psychiatry and Psychiatric Epidemiology, 36*, 287–293.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401. doi:10.1177/014662167700100306

Radloff, L. S. (1991). The use of the Center for Epidemiologic Studies Depression-Scale in adolescents and young adults. *Journal of Youth and Adolescents, 20*, 149-166. doi:10.1007/BF01537606

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163. doi:10.2307/271063

Roberts, R., E. Andrews, J. A., Lewinsohn, P. M., & Hops, H.(1990).Assessment of Depression in Adolescents Using the Center for Epidemiologic Studies Depression Scale.

Psychological Assessment, 2, 122-128.

Roberts, R. E., & Vernon, S. W. (1983). The Center for Epidemiologic Studies Depression Scale: its use in a community sample. *American Journal of Psychiatry, 140,* 41-46.

Rozario, P. A., & Menon, N.(2010).An examination of the measurement adequacy of the CES-D among African American women family caregivers. *Psychiatry Research, 179,* 107–112. doi:10.1016/j.psychres.2010.06.022

Rivera-Medina, C. L., Caraballo, J. N., Rodrıguez-Cordero, E. R., Bernal, G., & Davila-Marrero, E. (2010). Factor Structure of the CES-D and Measurement Invariance across Gen for Low-Income Puerto Ricans in a Probability Sample. *Journal of Consulting and Clinical Psychology, 78*, 398–408. doi: 10.1037/a0019054.

Ryder, A. G., Bean, G., & Dion, K. L. (2000). Caregiver responses to symptoms of first-onset psychosis: A comparative study of Chinese- and Euro-Canadian families. *Transcultural Psychiatry, 37*, 255–265. doi:10.1177/136346150003700207

Ryder, A. G., Yang, J., Zhu, X., Yao, S., Yi, J., Heine, S. J., & Bagby, R. M.(2008). The Cultural Shaping of Depression: Somatic Symptoms in China, Psychological Symptoms in North America? *Journal of Abnormal Psychology, 117*, 300–313. doi 0.1037/0021-843X.117.2.300

Schroevers, M. J., Sandersman, R., van Sonderen, E., Ranchor, A. V. (2000). The evaluation of the Center for Epidemiological Studies-Depression (CES-D) scale: depressed and positive affect in cancer patients and health reference subjects. *Quality of Life Research , 9*, 1015–1029.

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology, 62*, 123–146. doi:10.1002/jclp.20213

Tepper, P., Liu, X., Guo, C., Zhai, J., Liu, T., & Li, C. (2008). Depressive symptoms in

Chinese children and adolescents: Parent, teacher, and self reports. *Journal of Affective Disorders, 111*, 291–298. doi:10.1016/j.jad.2008.03.013

Thorson, J., & Powell, F. (1993). The CES-D: Four of five factors? *Bulletin of Psychonomic Society, 31*, 577–578.

Yen, S., Robins, C. J., & Lin, N. (2000). A cross-cultural comparison of depressive symptom manifestation: China and the United States. *Journal of Consulting and Clinical Psychology, 68*, 993–999. doi:10.1037//0022-006X.68.6.993

Ying, Y. (1988). Depressive symptomatology among Chinese-Americans as measured by the CES-D. *Journal of Clinical Psychology, 44*, 739-746. doi:10.1002/1097-4679(198809)44:5

Ying, Y., Lee, P. A., Tsai, J. L., Yeh, Y-Y., & Huang, J. S. (2000). The conception of depression in Chinese American College Students.*Cultural Diversity & Ethnic Minority Psychology, 6*, 183–195. doi: 10.1037/1099-9809.6.2.183

Zhang, B., Fokkema, M., Cuijpers, P., Li, J., Smits, N., & Beekman, A. (2011). Measurement invariance of the center for epidemiological studies depression scale (CES-D) among chinese and dutch elderly. *BMC Medical Research Methodology, 11*, 74. doi:10.1186/1471-2288-11-74

Zhang, J., Wu, Z., Fang, G., Li, J., Han, B., & Chen, Z.(2010). Development of the Chinese age norms of CES-D in urban area. *Chinese Mental Health Journal, 24*, 139-143.(In Chinese).

Table1.

Item Mapping for Tested Models

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

Note. DA = Depressed affect, IP = Interpersonal problem, PA = Positive affect, SC = Somatic complaints. M1 = Radloff's original four-factor model; M2 = Two factor model indentified in Cheung and Bagley' study (1998), in which DA + PA + SC as a factor and IP; M3 = another 2-factor model in which combined all negative items into a independent factor and 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.

Table 2.

Factor Loadings for EFA in sample 1.

Item content	SC	DA	PA
1. Bothered by things.	**.698**	.170	-.015
2. Appetite was poor.	**.481**	.091	.096
3. Can't shake off the blues.	**.664**	.249	.179
4. Just as good as others.	-.138	.127	**.604**
5. Trouble concentrating.	**.666**	.104	.033
6. Felt depressed.	**.673**	.365	.176
7. Everything was an effort.	**.594**	.272	.216
8. Hopeful about the future.	.111	.064	**.713**
9. Life has been a failure.	**.473**	.378	.376
10. Fearful.	**.562**	.388	.088
11. Sleep was restless.	**.454**	.223	.035
12. Happy.	.271	.136	**.694**
13. Talked less than usual.	.156	**.514**	.165
14. Felt lonely.	.363	**.613**	.180
15. People were unfriendly.	.167	**.753**	.172
16. Enjoyed life.	.223	.139	**.732**
17. Had crying spells.	.260	**.493**	-.150
18. Felt sad.	**.559**	**.503**	.059
19. People disliked me.	.219	**.767**	.155
20. Could not get going.	.236	**.616**	.216
Explained Variance	19.89%	16.95%	11.74%

Note. DA = Depressed affect, PA = Positive affect, SC = Somatic complaints.

Table 3

Goodness-of-Fit Indices and Model Comparisons for Tested Models

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

Note: CFI = Comparative Fit Index; TLI = Tucker-Lewis 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 indentified in Cheung and Bagley' study (1998), in which DA + PA + SC as a factor and IP; M3 = another 2-factor model in which combined all negative items 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 M 4 and allow item 9 and item 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 item 9 and item 18.

Table 4

Goodness-of-Fit Indices and Model Comparisons for Measurement Invariance Models

Model	WLSMVχ^2	*df*	TLI	CFI	RMSEA (90% CI)		$\Delta\chi^2$	Δdf	ΔTLI	ΔCFI
A-Configural invariance	2740.835**	264	.958	.963	.061 [.059 -, .063]					
B-Weak invariance	2849.063**	279	.958	.962	.060 [.058 - .062]	B vs. A	169.341**	15	.000	-.001
C-Scalar invariance	3233.528**	312	.957	.958	.061 [.059 - .063]	C vs. B	466.026**	33	-.001	-.004
D-Error Variance invariance	2942.618**	330	.961	.964	.056 [.054 - .058]	D vs. C	91.854**	18	+.004	+.006
E-Variances-covariances invariance	2520.412**	318	.967	.969	.052 [.050 - .054]	E vs. C	14.574*	6	+.006	+.005
F-Latent mean invariance	2512.446**	321	.968	.969	.052 [.050 - .054]	F vs. E	55.978**	3	+.001	.000

Note: * $p < .05$, ** $p < .01$; CFI = Comparative Fit Index; TLI = Tucker-Lewis index; RMSEA = Root Mean Square Error of a Approximation; CI = Confidence interval. $\Delta\chi^2$= Change in χ^2 relative to the preceding model; Δdf = Change in degree of freedom relative to the preceding model; ΔCFI = Change in comparative fit index relative to the preceding model; ΔTLI = 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.