FACTORIAL VALIDITY OF THE CENTER FOR EPIDEMIOLOGIC STUDIES-DEPRESSION (CES-D) SCALE IN MILITARY PEACEKEEPERS

Jennifer A. Boisvert, M.A., Donald R. McCreary, Ph.D., Kristi D. Wright, M.A., and Gordon J. G. Asmundson, Ph.D.

Despite widespread use of the Center for Epidemiologic Studies Depression Scale [CES-D], there are no investigations that examine its factor structure in a military sample. Separate confirmatory factor analyses were performed on responses to the CES-D obtained from 102 female and 102 male Canadian military peacekeepers in order to compare the fit of a four-factor intercorrelated (lower-order) model to a four-factor hierarchical (higher-order) model. The intercorrelated and hierarchical models fit the data well for both women and men, with hierarchical models fitting the data slightly better for women than men. These findings suggest that, for military women and men, the CES-D can be used to measure a set of distinct but interrelated depressive symptoms as well as a global construct of depression. Implications and future directions are discussed. Depression and Anxiety 17:19–25, 2003. © 2003 Wiley-Liss, Inc.

Key words: CES-D; depression, factorial validity; military; peacekeepers

INTRODUCTION

Depression has been found to be one of the most commonly diagnosed clinical disorders in community [e.g., Aalto-Setaelae et al., 2001; Kessler et al., 1994] and primary care settings [e.g., Blacker and Clare, 1987; Blacker et al., 1997; Katon and Schulberg, 1992]. Depression has also been identified as a predominant psychiatric complaint amongst military personnel [e.g., The Iowa Persian Gulf Study Group, 1997; Slusarcick et al., 1997; Winfield and Lafferty, 1997]. In military personnel, it is a significant concern given close associations with suicide, the third leading cause of death in military populations [Holmes et al., 1998; Sentell et al., 1997], and with the unique experiences that characterize military operations (e.g., exposure to novel environments, combat-related stressors, and atypical sleep schedules) [Ferrer et al., 1995; Sutker et al., 1994]. In order to better understand the nature of depression in military personnel and to improve associated assessment and treatment strategies, it is necessary to have a measure that provides a reliable and valid representation of the construct. One depression measure that has been used consistently throughout the epidemiological and psychological literature is the Center for Epidemiologic Studies Depression Scale (CES-D) [Radloff, 1977].

The Center for Epidemiologic Studies Depression Scale [CES-D; Radloff, 1977] is a standard screening instrument for depression in nonclinical populations and is “...designed to measure current level of depressive symptomatology, with emphasis on the affective component, depressed mood” (p 385). This scale taps into dimensions of depressive affect and is commonly used to determine a total score of depression. The CES-D has been widely used to assess depressive symptoms in community populations of varying age, sex, ethnicity, and language [e.g., Callahan and Wolinsky, 1994; Clark et al., 1981;...
Radloff [1977] asserted that the CES-D, to be considered a useful tool for epidemiologic research, “...must have adequate reliability and validity and a similar factor structure within each subgroup of the population” (emphasis original) (p 398). In other words, the CES-D factor structure must have adequate validity in order to ensure unbiased outcomes; that is, free of measurement bias within and across populations [e.g., Cole et al., 2000; Dean et al., 1994; Dean and Salem, 1998; Stommel et al., 1993]. Since its development, the CES-D scale has been subject to numerous factor analytic investigations in an attempt to validate its psychometric properties and applications.

Radloff [1977], in an initial exploratory factor analysis (EFA), identified and conceptualized four factors as independent subscales: 1) Depressed Affect (e.g., blues, depressed, lonely); 2) Positive Affect (e.g., good, hopeful, happy); 3) Somatic and Retarded Activity (e.g., bothered, sleep, get going); and 4) Interpersonal (e.g., unfriendly, dislike). Although several EFA investigations have been conducted [e.g., Beals et al., 1991; Callahan and Wolinsky, 1994; Clark et al., 1981; Devins et al., 1988; Joseph and Lewis, 1995; McCallion and Kolomer, 2000; Thorson and Powell, 1993], uncertainty remains regarding the CES-D's precise factor solution. Factor structure has varied between a two-factor [Edman et al., 1999; Manson et al., 1990], a three-factor [Beals et al., 1991; Kuo, 1984; Stroup-Benham et al., 1992; Ying, 1988], a five-factor [Thorson and Powell, 1993], and a seven-factor solution [Callahan and Wolinsky, 1994]. The variability of factor structure reported has been attributed to differences in demographic variables of study populations [e.g., Callahan and Wolinsky, 1994; Edman et al., 1999; Manson et al., 1990]. Results from other investigations have been consistent with Radloff's [1977] original four-factor model [e.g., Clark et al., 1981; Golding and Aneshensel, 1989; McCallion and Kolomer, 2000; Roberts et al., 1990; Stommel et al., 1993], although specific items loading on each factor have varied.

Other investigators have used confirmatory factor analysis (CFA) to test alternative models of the CES-D's four-factor solution: 1) four-factor intercorrelated (lower-order) model and 2) four-factor hierarchical (higher-order) model [e.g., Hertzog et al., 1990; Knight et al., 1997; Rhee et al., 1999]. Knight et al. [1997], in a community sample of women, found that both four-factor intercorrelated and hierarchical models provided a good fit. Hertzog et al. [1990] found comparable results in a sample of older adults. Sheehan et al. [1995], in a sample of patients suffering from rheumatoid arthritis (RA), examined four alternative models: 1) one-factor model; 2) three-factor model (depressed affect/positive affect, somatic, interpersonal); 3) four-factor intercorrelated model; and 4) four-factor hierarchical model. Akin to Knight et al. [1997], Sheehan et al. [1995] found that both four-factor intercorrelated and hierarchical models provided the best fit; however, the item content of the four factors were somewhat different than that proposed by Radloff [1977].

Most recently, Rhee et al. [1999] reexamined the CES-D factor structure in RA patients. In addition to the four models examined by Sheehan et al. [1995], Rhee et al. [1999] also explored an additional three-factor model (depressed affect/somatic, positive affect, and interpersonal) that was supported by results from a previous study in a RA sample. Furthermore, all of the three- and four-factor models were tested with both the Radloff item allocation as well as the Sheehan item allocation [see Rhee et al., 1999 for item allocation]. The results suggested that both the four-factor intercorrelated and the four-factor hierarchical model (with the Radloff item allocation) were equivalent in providing the best fit.

While the factor structure of the CES-D has been studied in various community samples, its factorial validity has not yet been explored in a military sample. Since military personnel and civilians may differ in the experiences that lead to depression (e.g., exposure to combat-related stressors), it cannot be assumed that the factor structure of the CES-D is equivalent in each. Thus, the purpose of the present study was to establish the factor structure of the CES-D in a sample of Canadian peacekeepers. Also, because the factor structure of the CES-D in female and male military personnel may not be the same [i.e., there may be sex differences in the factor intercorrelations; McCrea et al., 1998], we performed separate CFA on responses to the CES-D obtained from each of the sexes. The following models were tested: 1) four-factor intercorrelated model and 2) four-factor hierarchical model [e.g., Hertzog et al., 1990; Knight et al., 1997; Rhee et al., 1999; Sheehan et al., 1995].

METHOD

PARTICIPANTS AND PROCEDURE

Participants were selected from a larger sample of 1968 veterans from regular and reserve duty forces of the Canadian military who, as part of an anonymous health status assessment conducted by Veterans' Affairs Canada in the Fall of 1999, voluntarily completed a self-administered battery of questionnaires. There were 107 female participants in the sample and, of these, 102 provided complete information on the measures pertinent to this study. For comparative purposes, 102 male participants were selected at random from the larger sample of those who provided complete information on the measures if interest. For women, the mean age was 42.1 years (SD=9.0). Approximately 60% (n=61) were married/common-law, 95% (n=95) had completed high school, 90% (n=91) were noncommissioned officers, and 56% (n=50) had been medically released from the Canadian Forces.
For men, the mean age was 45.6 years (SD=12.4). Approximately 84% \((n=85)\) were married/common-law, 84.5% \((n=85)\) completed high school, 66% \((n=64)\) noncommissioned officers, and 38% \((n=27)\) had been medically released from the Canadian Forces. Significant sex differences \((P<.05)\) were found with respect to the following demographic variables: age, marital status, educational status, military rank, and medical status.

**MEASURES**

The CES-D [Radloff, 1977] was administered to participants to evaluate depressive symptoms. The CES-D is a 20-item self-report instrument that asks respondents to describe their mood over the past week on a 3-point frequency scale \([0=\text{rarely or none of the time (less than 1 day)}, 1=\text{some or a little of the time (1–2 days)}, \text{and 3=most or all of the time (5–7 days)}]\). While not a diagnostic tool, the CES-D does yield a total score that is indicative of degree of depression. Scores of 16 or greater are considered to represent possibly clinically significant depression [Barnes and Prosen, 1984; Radloff, 1977]. Radloff [1977] reports an adequate level of internal consistency \((\alpha>.84)\) and test–retest reliability ranging from 0.49 (12 months) to 0.67 (4 weeks). Lower reliabilities for longer test-retest intervals have been attributed to, amongst other things, the greater likelihood of cyclic changes in depressive symptoms and depressive reaction to life events [Radloff, 1977]. The CES-D has demonstrated good convergent validity with other measures of depressive symptoms [e.g., \(r>.50\) with the Hamilton Rating Scale for Depression; Devins and Orme, 1985]. Table 1 shows descriptive statistics for the CES-D total score as well as the depressed affect, positive affect, somatic complaints, and interpersonal distress subscales for men and women in our sample.

**RESULTS**

The structural modelling program EQS [Bentler, 1993], along with the procedures outlined by Byrne [1996], were used to conduct a series of CFAs of the CES-D. For all analyses, a maximum-likelihood solution was used. In one set of CFAs, all four CES-D factors were allowed to intercorrelate. Because the goal was to determine whether the proposed four-factor structure fit the observed data, and not whether individual CES-D items loaded onto specific latent factors, no correlated residuals or cross-loadings were allowed. To test whether a four-factor hierarchical structure provided a good or better fit to the observed data, the factor structure from the intercorrelated model was altered by removing the correlational paths among latent factors and adding directional paths from a higher order CES-D factor to each latent factor. Each CFA model (i.e., the intercorrelated and hierarchical models) was tested separately for women and men. Our sample size was insufficient to permit separate item-level analyses (i.e., analyses using individual CES-D items) for women and men.

The items from each of the four CES-D subscales were randomly assigned to one of two indicators per subscale. These indicators then were used to create four latent factors representing those subscales [see Byrne, 1996]. As recommended by Brown and Cudeck [1993] and Hu and Bentler [1998], multiple fit indices were used to determine how well the proposed factor structure fit the observed data. The fit indices included a) \(\chi^2\) (values should not be significant, but in larger samples this is often not feasible); b) \(\chi^2/df\) ratio (values should be <2.0); c) Adjusted Goodness-of-Fit Index (AGFI; values should be >0.80); d) Comparative Fit Index (CFI; values should be >0.95); e) Root Mean Square Error of Approximation (RMSEA; values should be around .05); and f) the Standardised Root Mean Square Residual (SRMR; values should be approximately .08) [Brown and Cudeck, 1993; Hu and Bentler, 1998; Marsh et al., 1988]. To determine whether the higher-order model provided a better fit to the observed data than the intercorrelated model, a \(\chi^2\) difference test was used to compare models separately for women and men.

**Men's models.** The first CFA explored the degree to which intercorrelated factor structure explained the observed variability in CES-D scores. An initial analysis revealed the presence of two extreme multivariate outliers. These two cases were removed and all further analyses were conducted using the remaining 100 cases. The results of the intercorrelated model CFA showed that all factor loadings and variances were significant \((P<.01)\) and all but one item residual (i.e., Positive Affect 2) was significantly different from zero \((P<.01)\). Table 2 shows the fit indices for the men's intercorrelated model. As can be seen, the \(\chi^2\) statistic is not significant and all other fit indices are within their optimal range. Table 3 shows the correlations among the four CES-D latent factors. The final model is shown in Figure 1.

The second CFA tested a hierarchical model. Thus, all the factor intercorrelations were restricted as directional paths and the CFA was rerun. The test of the men's intercorrelated model CFA showed that all factor loadings and variances were significant \((P<.01)\)

**TABLE 1. CES-D total and subscale scores**

<table>
<thead>
<tr>
<th>Score*</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>CES-D total</td>
<td>12.0 (10.5)</td>
<td>15.2 (14.3)</td>
</tr>
<tr>
<td>DA</td>
<td>2.8 (4.3)</td>
<td>4.8 (5.6)</td>
</tr>
<tr>
<td>PA</td>
<td>4.9 (3.8)</td>
<td>4.6 (4.0)</td>
</tr>
<tr>
<td>SC</td>
<td>3.8 (4.1)</td>
<td>5.2 (5.2)</td>
</tr>
<tr>
<td>IR</td>
<td>0.4 (1.0)</td>
<td>0.7 (1.3)</td>
</tr>
</tbody>
</table>

*Values are expressed as mean SD.

DA, depressive affect; PA, positive affect (items reverse-scored); SC, somatic complaints; IR, interpersonal relationships.
and all but one item residual (i.e., Interpersonal Relations 2) was significantly different from zero ($P < .01$). In addition, all the directional paths between the CES-D subscales and the higher-order factor were significant ($P < .01$). As Table 2 shows, while the CFI and AGFI were within their optimal range, the other fit statistics were out of range. An examination of the disturbance terms shows that the higher-order factor explained between 26% and 90% of the variability in the first order factors (i.e., 1.00 minus disturbance terms squared). The hierarchical model is shown in Figure 2. The fact that the intercorrelated model provides a better fit to the data was shown empirically using a $\chi^2$ difference test (see Table 2).

**Women’s models.** As with the men’s models, an initial analysis revealed the presence of two multivariate outliers. These cases were removed from all further analyses. The test of the women’s intercorrelated model showed that all factor loadings and variances were significant ($P < .01$) and all but one item residual (i.e., Interpersonal Relations 2) was significantly different from zero ($P < .01$). As Table 2 shows, while the CFI and AGFI were within their optimal range, the other fit statistics were out of range. An examination of the disturbance terms shows that the higher-order factor explained between 26% and 90% of the variability in the first order factors (i.e., 1.00 minus disturbance term squared). The hierarchical model is shown in Figure 2. The fact that the intercorrelated model provides a better fit to the data was shown empirically using a $\chi^2$ difference test (see Table 2).

**TABLE 2. Fit indices for intercorrelated and hierarchical model confirmatory factor analyses**

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$P$</th>
<th>$\chi^2/df$</th>
<th>AGFI</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IM</td>
<td>19.15</td>
<td>14</td>
<td>0.16</td>
<td>1.37</td>
<td>0.89</td>
<td>0.99</td>
<td>0.06</td>
<td>0.09</td>
</tr>
<tr>
<td>HM</td>
<td>38.17</td>
<td>18</td>
<td>0.01</td>
<td>2.12</td>
<td>0.84</td>
<td>0.96</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>$\chi^2$ difference</td>
<td>19.02</td>
<td>4</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IM</td>
<td>21.76</td>
<td>14</td>
<td>0.08</td>
<td>1.55</td>
<td>0.88</td>
<td>0.99</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>HM</td>
<td>25.51</td>
<td>17</td>
<td>0.08</td>
<td>1.50</td>
<td>0.88</td>
<td>0.99</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>$\chi^2$ difference</td>
<td>3.75</td>
<td>3</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*IM, intercorrelated model; HM, hierarchical model; AGFI, Adjusted goodness of fit test; CFI, comparative fit index; RMSEA, root mean square error of approximation; SRMR, squared root mean residual.*

**TABLE 3. Latent factor intercorrelations**

<table>
<thead>
<tr>
<th></th>
<th>DA</th>
<th>PA</th>
<th>SC</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>DA</td>
<td>0.57</td>
<td>0.88</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>0.78</td>
<td>0.37</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>0.96</td>
<td>0.74</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>0.69</td>
<td>0.45</td>
<td>0.65</td>
<td></td>
</tr>
</tbody>
</table>

*Values expressed reflect men above the median and women below the median.

DA, depressive affect; PA, positive affect (items reverse-scored); SC, somatic complaints; IR, interpersonal relationships.*

![Figure 1. Final confirmatory factor analytic model for intercorrelated model. Rectangles are measured variables and large circles are latent factors. Numbers associated with single headed arrows running from circles to rectangles are factor loadings. All factor loadings are standardised and statistically significant ($P < .01$), except where indicated (ns). Small sourceless arrows represent item residuals and all are statistically significant ($P < .01$). Women’s loadings are to the left of the back slash and men’s loadings are to its right.](image-url)

Disturbance terms (i.e., Depressive Affect, Somatic Complaints) were significantly different from zero ($P < .01$). In addition, all the directional paths between the CES-D subscales and the higher-order factor were significant ($P < .01$). As Table 2 shows, all fit statistics were within their optimal range. Examination of the disturbance terms shows that the higher-order factor explained between 47% and 100% of the variability in
the first order factors. The hierarchical model is shown in Figure 2. The \( \chi^2 \) difference test showed that both the intercorrelated and hierarchical models provided a similar fit to the observed data (see Table 2).

**DISCUSSION**

The purpose of the present study was to establish the factorial validity of the CES-D in a military sample using CFA. This is a statistical procedure that is used to assess the factorial validity of a measure when there is an existing model against which to compare data. We compared CES-D data from women and men separately against the four-factor intercorrelated and hierarchical models. Both intercorrelated and hierarchical models fit the data for women and men. The magnitude of the differences in model fit was slight, suggesting that the models fit similarly for women and men. Our results are relatively consistent with Radloff's [1977] four-factor structure of the CES-D. Study findings further corroborate results from previous CFA investigations [e.g., Hertzog et al., 1990; Knight et al., 1997; Rhee et al., 1999; Sheehan et al., 1995] in evaluating the appropriateness of intercorrelated and hierarchical model representation of the CES-D. The stability of the four-factor intercorrelated model implies that the responses to the CES-D may be conceptualized as an amalgam of four related but distinct depression subscales. Yet, the fit of the four-factor hierarchical model suggests that the CES-D can also be conceptualized as a global construct of depression.

For men, the fit statistics for the intercorrelated model were good and all but one item residual (Positive Affect 2, see Fig. 1) were significantly different from zero. These results suggest that for men, the four-factor intercorrelated model of the CES-D appears to be conceptually valid. Only the CFI and AGFI fit statistics were within their optimal range for the hierarchical model. Furthermore, the results from the \( \chi^2 \) difference test suggested that the intercorrelated model provided better fit to the data than the hierarchical model.

For women, the fit statistics for the intercorrelated model were fairly good. All but one item residual (Interpersonal Relations 2, see Fig. 1) was significantly different from zero. These results suggest that the four-factor intercorrelated model of the CES-D is conceptually valid for women. The hierarchical model provided a good fit to the data as well. All but one item residual (Interpersonal Relations 2, see Fig. 2) and two latent factor disturbance items (i.e., Depressed Affect and Somatic Complaints) was significantly different from zero. The \( \chi^2 \) difference test showed that both models provided similar fit to the observed data.

Although we conducted separate analyses for women and men, the two groups were not compared against each other. However, as evident in factor analytic investigations of the CES-D [e.g., Roberts et al., 1990] as well as other measures [e.g., Wright et al., 2001], the factor structure of item responses can differ between men and women. These differences can be attributed to various factors, including the possibility that women and men have different response patterns to different items on the CES-D [also see Stommel et al., 1993] and that their experience with symptoms differs. Indeed, in the military setting it has been observed that women have greater difficulty in coping with combat-related stressors and have more severe depressive symptoms than do men [Breslau et al., 1997, 2000; Wagner et al., 2000]. Given these possibilities, researchers and clinicians working with military personnel will need to consider which scoring method (i.e., total score or four subscale scores) will maximize information when screening for depressive symptoms in women and men. Given the minor differences in factor structure observed in the present study, we suggest assessing both total and subscale scores.

One potential limitation of our study is that we did not conduct a true CFA in the sense that we examined the factor structure of the CES-D at the subscale level rather than at the item level. Therefore, it could be argued that we were presumptuous in using only Radloff's [1977] four-factor solution as this solution has not been consistently reported across all samples. In examining the CES-D factor structure at the subscale level, we made the assumption that the subscales had good internal consistency and validity. Notwithstanding, our decision to use this approach was based on the results of the aforementioned CFA investigations [e.g., Rhee et al., 1999] that confirmed Radloff's [1977] four-factor structure of the CES-D. A second limitation is that, given the relatively small sample of female participants, we were unable assess CES-D factor structure between the sexes as a function of exposure to combat (e.g., as a result of being deployed to a combat theatre).

This investigation is the first to examine the CES-D factor structure in a military sample and, as such, provides information relevant to utilization of the CES-D in future investigations of depression in military personnel. Both the total score and the four subscale scores of the CES-D were shown to provide factorially valid information in military personnel. Consequently, application of the CES-D in symptom assessment will provide useful information regarding overall depressive symptoms as well as information specific to positive and depressed affect, somatic complaints, and interpersonal relationships. The richness of the information provided by the CES-D subscales may prove particularly beneficial when assessing depression in those personnel who are prone to have elevated somatic complaints (e.g., those deployed to combat theatres) and, possibly, for tailoring interventions to most appropriately address symptoms profiles. The measure may also prove valuable in evaluating the outcome of treatment provided to military personnel with symptoms of depression. Whether these potential benefits meet or exceed other
short and psychometrically sound measures of depression remains to be evaluated.

Acknowledgments. We thank Veteran’s Affairs Canada for providing access to the data contained within this article. The information used was extracted from a survey conducted in the Fall 1999 by the Review of Veterans’ Care Needs project, Veteran’s Affairs Canada. We also thank Dr. W. Andrew Harrell for his feedback on earlier versions of this manuscript.

REFERENCES


