Psychometric Properties of the CES-D-10 in a Psychiatric Sample

Thröstur Björgvinsson[1], Sarah J. Kertz[1], Joe S. Bigda-Peyton[1], Katrina L. McCoy[1], and Idan M. Aderka[1]

Abstract

The 10-item Center for the Epidemiological Studies of Depression Short Form (CES-D-10) is a widely used measure to screen for depression in primary care settings. The 10-item measure has demonstrated strong psychometric properties, including predictive accuracy and high correlations with the original 20-item version, in community populations. However, clinical utility and psychometric properties have yet to be assessed in an acutely symptomatic psychiatric population. This study examined the psychometric properties of the CES-D-10 in a sample of 755 patients enrolled in a psychiatric partial hospital program. Participants completed a diagnostic interview and a battery of self-report measures on admission and discharge. Exploratory factor analysis and confirmatory factor analysis suggested that a one-factor structure provided a good fit to the data. High item–total correlations indicated high internal consistency, and the CES-D-10 demonstrated both convergent validity and divergent validity. Previously suggested cutoff scores of 8 and 10 resulted in good sensitivity (.91 and .89, respectively) but poor specificity (.35 and .47). These data suggest that although the CES-D-10 has generally strong psychometric properties in this psychiatric sample, the measure should be primarily used to assess depression symptom severity rather than as a diagnostic screening tool.

Major depressive disorder (MDD) is one of the most common mental health disorders, with an estimated lifetime prevalence rate of 17.0% for adults in the United States (Kessler et al., 2003; Richards, 2011). MDD often follows a chronic course and is associated with significant functional impairment in relationships and occupational functioning (Kessler et al., 2003). MDD is also associated with an increased risk of suicide, hospitalization, and increased health care utilization (Howland, 1993), resulting in an economic burden of $83.1 billion on the United States in 2000 (Greenberg et al., 2003).

Given the prevalence, associated impairment, and economic impact of MDD, accurate assessment of the presence and severity of depressive symptoms is imperative. Researchers and clinicians have developed several self-report measures used in primary care settings (Helmreich et al., 2011; Mitchell, Rao, & Vaze, 2011; Sharp & Lipsky, 2002). The Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977) is one such measure; it has evidenced strong psychometric properties in assessing symptoms of depression and detecting a depression diagnosis (Radloff, 1977). The original version of the CES-D was shortened from 20 to 10 items to improve clinical utility and ease of scoring (Andersen, Malmgren, Carter, & Patrick,

1994). This revised version (CES-D-10) has demonstrated strong psychometric properties, indicated by good reliability and construct validity in older adults (Irwin, Artin, & Oxman, 1999), multicultural populations (Cheng, Chan, & Fung, 2006), and adolescents (Bradley, Bagnell, & Brannen, 2010). Moreover, the CES-D-10 has been shown to adequately screen for hopelessness and suicidality in community samples (Y. Cheung, Liu, & Yip, 2007).

The psychometric properties of the CES-D-10 in adult clinical samples are not well understood. Most studies examining the CES-D-10 have focused on the general population, or other specific subgroups (e.g., geriatric populations, HIV-infected veterans, etc.) reviewed above. Although a small number of studies have used the CES-D-10 to screen for depression among adult psychiatric populations (e.g., Kilbourne, Justice, Rabeneck, Rodriguez-Barradas, & Weissman, 2001), the utility of the CES-D-10 in psychiatric settings is not well established. Based on the

[1]McLean Hospital/Harvard Medical School, Belmont, MA, USA

Corresponding Author:
Thröstur Björgvinsson, McLean Hospital, 115 Mill St., Belmont, MA 02478, USA.
Email: tbjorgvinsson@mclean.harvard.edu

small body of literature examining the utility of CES-D-10 among psychiatric populations, there is evidence that it may be an adequate screener for depression among psychiatric samples (Irwin et al., 1999). In fact, the CES-D-10 may be a particularly accurate depression screener when comorbid diagnoses are present (Cheng & Chan, 2008). However, studies examining the utility of CES-D-10 among psychiatric samples are few and focus on specific populations. Despite the lack of evidence that CES-D-10 is an adequate screening tool among general psychiatric populations, its feasibility makes it an attractive and more frequently used tool in such settings (Nishiyama, Ozaki, & Iwata, 2009). However, in the absence of more sound empirical examination, it cannot be assumed that the measure performs equally well in psychiatric settings where rates of severe depression have been found to be as high as 37% (Ciro et al., 2012). Given the severity and heterogeneity of symptom presentations in general clinical settings, examining how the measure performs in a diagnostically heterogeneous, acutely symptomatic, psychiatric sample would provide a more rigorous test for the CES-D-10. Such information could address a large gap in the literature on the CES-D-10 and inform its use in psychiatric settings.

Analyses of the underlying factor structure of the CES-D-10 have been generally inconsistent. For example, several studies have suggested that a two-factor structure, including both a 2-item positive affect factor and an 8-item negative affect factor, provided the best fit to the data in an adolescent sample (Bradley et al., 2010; Lee & Chokkanathan, 2008; Zhang et al., 2012); however, in a sample of older Chinese adults a three-factor structure emerged (Cheng et al., 2006). Another study found that a single-factor model provided a good fit to the full-length version of the CES-D and two shortened versions (Carpenter et al., 1998). These contradictory findings may be due in part to the use of specific subgroups of individuals (e.g., older Chinese adults) that may limit generalizability. Discrepant findings might also be explained by a failure to account for the potential impact of method effects on resultant factor structures, an issue that has thus far been unstudied in relation to the CES-D-10. For instance, some evidence suggests that reverse-scored items may form a distinct factor based on reverse-phrasing rather than content (Weeks et al., 2005). Specifically, two of the CES-D-10 items are worded in the reverse ("I was happy" and "I felt hopeful about the future") so that higher scores indicate less depression. These two items formed the positive affect factor identified in the Lee and Chokkanathan (2008) and Cheng et al. (2006) studies. Of note, both Cheng et al. (2006) and Lee and Chokkanathan (2008) reported that the positive affect factor appeared to have low construct validity. Furthermore, factors with fewer than three items are undesirable, given that they are

generally weak and unstable (Costello & Osborne, 2005). This factor structure raises questions about whether or not the factor is theoretically and practically meaningful (i.e., does this factor represent a relevant construct and does it have clinical utility) or whether it simply reflects a method effect related to the reversed wording of the items (see Brown, 2006, for a discussion). Allowing the error variances of the two reverse-worded items using a confirmatory factor analysis (CFA) would provide a formal test of this hypothesis, and such work is needed.

The current study evaluated the psychometric properties of the CES-D-10 in a sample of diagnostically heterogeneous, acutely symptomatic, partial hospital patients. The aims of the study were to examine several characteristics of the CES-D-10, including its factor structure, internal consistency, convergent and divergent validity, and functionality as a screening tool for depression. Due to the lack of consensus regarding the factor structure of the CES-D-10 and a paucity of literature in clinical samples, we first conducted an exploratory factor analysis (EFA), followed by a CFA accounting for method effects. We predicted that the measure would show good sensitivity and adequate specificity in identifying participants with a current major depressive episode.

Method

Participants and Procedure

Participants were 755 patients seeking treatment at the Behavioral Health Partial Program, a partial hospitalization program using individual and group cognitive behavioral therapy to treat a variety of Axis I and II personality disorders. Half of the patients (47.5%) were referred by outpatient treatment providers for an increased level of care whereas the other half (52.5%) were referred from inpatient hospitalization. Demographic characteristics are summarized in Table 1. Participants were diagnosed using the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). Diagnostic comorbidity was common, with almost half of the patients (49.0%; $n = 370$) meeting criteria for more than one Axis I disorder. See Table 1 for diagnostic characteristics.

The study was approved by the hospital's internal review board, and the participants were treated in accordance with the ethical guidelines of the American Psychological Association. All study participants provided written informed consent prior to the study. At admission participants completed the MINI, a demographics survey, and a battery of self-report measures; the self-report measures were completed again at discharge. Data were collected from July 2010 to November 2011. The MINI was administered by doctoral students in clinical psychology and predoctoral psychology interns. Students met for weekly

Table 1. Demographics and Diagnoses (*n* = 755).

	M (*SD*) or *N* (%)
Age	35.04 (13.57)
Gender	
Male	337 (44.6%)
Female	418 (55.4%)
Ethnicity[a]	
American Indian	5 (0.7%)
Asian	28 (3.7%)
African American	12 (1.6%)
Caucasian	646 (85.8%)
Caribbean Islander	2 (0.3%)
Latino/a	13 (1.7%)
Multiracial	21 (2.8%)
Choose not to respond/Do not know	26 (3.4%)
Education[a]	
Some high school	7 (0.9%)
High school graduate	292 (38.9%)
Some college	75 (10.0%)
College graduate	165 (22.0%)
Postcollege education	212 (28.2%)
Employment status[a]	
Employed	338 (44.8%)
Not employed	416 (55.2%)
Marital status[a]	
Never married	437 (58.1%)
Separated/divorced	105 (14.0%)
Widowed	6 (0.8%)
Married	182 (24.2%)
Living with partner	22 (2.9%)
Diagnosis[b]	
Major depressive disorder—Currently depressed	415 (55.0%)
Bipolar disorder—Currently depressed	98 (13.0%)
Major depressive disorder—Recurrent (not depressed)	109 (14.4%)
Bipolar disorder—Currently manic or mixed	61 (4.4%)
Psychotic disorder	49 (6.5%)
Generalized anxiety disorder	256 (33.9%)

a. Missing data ranges from 1 to 4, so total is less than 755.
b. Percentages exceed 100% due to comorbidity.

supervision with a psychology postdoctoral fellow. Those who administered the MINI did not have access to the CES-D-10 when they administered the MINI.

Measures

Miniature International Neuropsychiatric Interview. The MINI (Sheehan et al., 1998) is a structured interview assessing for Axis I symptoms as outlined by the *Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition* (*DSM-IV*; American Psychiatric Association, 1994). The MINI has demonstrated strong reliability with the Structured Clinical Interview for *DSM-IV*, with interrater reliabilities ranging from kappas of .89 to 1.0 (Sheehan et al., 1998).

Center for Epidemiological Studies Depression Scale-10. The CES-D-10 (Andersen et al., 1994) is a brief, widely used, self-report instrument assessing for depression over the past week. Responses are rated on a 4-point scale from 0 (*less than one day*) to 4 (*5-7 days*). The CES-D-10 has demonstrated adequate reliability and validity, with good internal consistency in this study (Cronbach's α = .89).

Behavior and Symptom Identification Scale (BASIS-24). The BASIS-24 (Eisen, Normand, Belanger, Spiro, & Esch, 2004) is a 24-item self-report measure that includes six subscales: Depression and Functioning, Interpersonal Relationships, Psychosis, Substance Abuse, Emotional Lability, and Self-harm. It has demonstrated good psychometric properties across inpatient, outpatient, residential, and partial hospital settings (Eisen et al., 2004). The Depression and Functioning subscales showed high reliability in this study with Cronbach's α = .89.

Penn State Worry Questionnaire–Abbreviated (PSWQ-A). The PSWQ-A (Hopko et al., 2003) is an abbreviated, 8-item measure designed to assess worry severity, derived from the original 16-item instrument (Meyer, Miller, Metzger, & Borkovec, 1990) with good reliability and validity (Hopko et al., 2003). Reliability in the present study was very high (Cronbach's α = .95).

Schwartz Outcome Scale (SOS). The SOS (M. A. Blais et al., 1999) is a well-validated and reliable measure assessing overall psychological well-being (M. R. Blais, Kehl-Fie, & Blias, 2008). Internal consistency of the SOS was high in the present study (Cronbach's α = .94).

Emotion Regulation Questionnaire (ERQ). The ERQ (Gross & John, 2003) is a 10-item self-report inventory assessing use of emotion regulation strategies, including reappraisal and behavioral suppression. Internal consistency of the reappraisal subscale was high (Cronbach's α = .80), while the suppression subscale was moderate (Cronbach's α = .73) in the present study.

Data Analysis

SPSS version 19.0 was used for all analyses other than the confirmatory factor analyses, which used LISREL version 8.80. As in many clinical samples, our data were nonnormally distributed (Shapiro–Wilk statistic = 0.97, *df* = 803, *p* < .001; skewness = −.31, *SE* = .09, kurtosis = −.87, *SE* = .17).

Therefore, we used adjusted χ^2 tests, nonparametric tests, and reported medians throughout the analysis.

The sample was randomly split into two subsamples for the exploratory principal components analysis and confirmatory factor analyses. The samples were then combined for the remaining analyses. We first conducted an EFA with principal components analysis. Given the nonnormality of our data, we used syntax for parallel analysis with raw data to determine the proper number of components to extract (O'Connor, 2000). Parallel analysis extracts eigenvalues from randomly generated data sets that parallel the parameters of the research data. The mean eigenvalues and those that correspond to the 95th percentile of the distribution of random data eigenvalues are then compared to those from the research data. Components are retained when the eigenvalue from the research data is greater than the randomly generated values.

We next conducted a CFA on the second half of the dataset. Because our data were nonnormal, we used the robust maximum likelihood estimation (MLM), which has been shown to perform well under conditions of nonnormal data (Brown, 2006), and the Satorra–Bentler scaled χ^2 (SB χ^2; Satorra & Bentler, 1994). We used the following goodness of fit indices: (a) the root mean square error of approximation (RMSEA), which should be less than .06 for an adequate model; (b) the comparative fit index (CFI); (c) the goodness-of-fit index (GFI), with values >.95 required for a well-fitting model for both CFI and GFI (Bentler & Bonett, 1980; Hu & Bentler, 1999); and (d) the standardized root mean square residual (SRMR) close to .08 or below (Brown, 2006).

Item characteristics, internal consistency, and convergent and divergent validity were assessed using Spearman's ρ, a nonparametric bivariate correlation estimate. We considered correlations <.40 low, .40 to .69 moderate, and >.69 high. We also used a Mann–Whitney test to compare CES-D-10 scores in participants who do and do not meet criteria for a current major depressive episode. A receiver operating characteristic (ROC) curve was used to examine the CES-D-10 compared to a diagnosis of a major depressive episode from the MINI. Sensitivity, specificity, and positive and negative predictive values were calculated for a range of cutoff scores. The ideal cutoff score was calculated by giving equal weight to sensitivity and specificity.

Results

Exploratory Factor Analysis

We used Sample 1 (n = 379) for the exploratory analysis. We first examined model assumptions. The Kaiser–Meyer–Olkin measure of sampling adequacy was acceptable, .91, and Bartlett's test of sphericity was significant, p < .001.

Taken together, this suggests that factor analysis is appropriate for these data. Results of the parallel analysis with 1,000 data sets suggested that a one-factor solution was best, given that only the first eigenvalue from the actual data (5.06) was greater than the corresponding 95th percentile (1.34) and mean (1.26) random data eigenvalues. We then conducted the principal components analysis specifying the extraction of one factor. The single factor explained 50.55% of the variance after extraction. The factor loadings ranged from .51 (Item 5) to .86 (Item 3). Because EFA does not allow for the modeling of method variance, we examined the influence of the effect of the reverse wording by using a CFA.

Confirmatory Factor Analysis

Based on the results from the EFA, we tested two models. The first examined the fit of a one-factor model, followed by a test of one-factor model with adjustments made according to error theory. Because we hypothesized that the two-factor model may be due to method effects associated with the reverse scoring of the two items on the "positive affect" factor, we therefore specified a unidimensional factor structure, with correlated error terms of Items 5 and 8 to account for unique variance shared by the items due to reverse wording.

Fit indices for the unidimensional model suggested that the model did not provide a good fit to the data, $SB \chi^2(df = 35) = 136.00$, GFI = .92, CFI = .97, RMSEA = .09, SRMR = .05, as evidenced by a high RMSEA and low GFI. Estimates from the standardized solution are presented in Table 2, with loadings ranging from .43 to .87. We next allowed the error terms of Items 5 and 8 to covary to account for method variance between the two reverse-worded items. Results of the revised model indicated that the revised model fit the data well, SB $\chi^2(df = 34) = 80.06$, GFI = .95, CFI = .99, RMSEA = .06, SRMR = .04. The covariance between the error terms of Items 5 and 8 was .23, and factor loadings are presented in Table 2. The χ^2 difference test was significant, $\Delta SB \chi^2(df = 1) = 55.94$, p < .01. The change in CFI (>.01) also suggests significant improvement in the model (G. W. Cheung & Rensvold, 2002).

Overall, these results suggest that a one-factor model, accounting for the unique variance between the two reverse-worded items, fits the data well and is significantly better than a model that does not account for variance between the two reverse-worded items. Furthermore, such a model appears to be conceptually meaningful and practically useful.

Item Characteristics and Internal Consistency

The overall median CES-D-10 score in this sample was 18.00, which is 3.5 times greater than the mean CES-D-10

Table 2. Item Characteristics and Factor Loadings for CES-D-10 Items.

Each item starts with the phrase: "During the PAST WEEK . . ."	*Mdn*	Corrected item–total correlation	EFA factor loading	CFA factor loading
1. I was bother by things that usually don't bother me.	1.0	.57	.64	.61
2. I had trouble keeping my mind on what I was doing.	2.0	.64	.72	.69
3. I felt depressed.	2.0	.79	.86	.87
4. I felt like everything I did was an effort.	2.0	.74	.84	.80
5. I felt hopeful about the future.[a]	2.0	.49	.51	.60
6. I felt fearful.	1.0	.58	.70	.59
7. My sleep was restless.	2.0	.44	.59	.44
8. I was happy.[a]	2.0	.60	.74	.67
9. I felt lonely.	2.0	.64	.66	.66
10. I could not get "going."	2.0	.72	.78	.80
Total score	18.0	—	—	—

Note. CES-D-10 = Center for the Epidemiological Studies of Depression Short Form (10 items); EFA = exploratory factor analysis; CFA = confirmatory factor analysis. Corrected item–total correlations are the correlation between the item and the total score without that item.
a. Indicates reverse-scored items.

Table 3. Convergent Validity: Spearman's Rho Correlations.

	1	2	3	4	5	6
1. CES-D-10	—	—	—	—	—	—
2. BASIS-DF	.86***	—	—	—	—	—
3. PSWQ-A	.46***	.49***	—	—	—	—
4. SOS	−.54***	−.47***	−.48***	—	—	—
5. BASIS-PS	.18***	.18***	.07*	−.01	—	—
6. BASIS-SA	.09*	.11**	.01	.04	.22***	—
7. ERQ	−.23***	−.26***	−.25***	.37***	.03	.06

Note. CES-D-10 = Center for the Epidemiological Studies of Depression Short Form (10 items); BASIS-DF = BASIS24–Depression and Functioning Subscale; PSWQ-A = Penn State Worry Questionnaire–Abbreviated; SOS = Schwartz Outcome Scale; BASIS-PS = BASIS24–Psychosis subscale; BASIS-SA = BASIS24–Substance Abuse subscale; ERQ = Emotion Regulation Questionnaire.
***p < .001. **p < .01.*p < .05.

score of 4.70 in the general population (Andersen et al., 1994). Medians of each item of the CES-D-10 and the total score are presented in Table 3. Item–*total* correlations ranged from moderate ("My sleep was restless"; ρ = .45, *p* < .01) to high ("I felt depressed"; ρ =.81, *p* < .01).

Convergent Validity

Convergent validity of the CES-D-10 was assessed by examining correlations with the BASIS-24–Depression and Functioning subscale, worry, and overall well-being (see Table 3). As expected, the CES-D-10 correlated positively and strongly with Depression and Functioning and moderately with worry. The CES-D-10 total score also showed moderate negative correlations with overall psychological well-being. We also found that those with a current major depressive episode scored significantly higher than those without a current episode, *U* = 46835.50, *Z* = −7.96, *p* ≤ .001, *Mdn* = 20.00 and *Mdn* = 14.00, respectively.

Divergent Validity

Divergent validity of the CES-D-10 was assessed by examining correlations with the BASIS-24–psychosis and substance abuse subscales, as well as the ERQ. As expected, results indicated that the CES-D-10 had very low correlations with these scales (see Table 3).

Cut-Scores, Sensitivity, and Specificity

An ROC analysis estimated the area under the curve at .81 (95% confidence interval = .78, .85). Cutoff scores of both 8 and 10 have been suggested for the CES-D-10 in nonclinical, older adult samples (Andersen et al., 1994). In the present sample, a cutoff score of 8 resulted in good sensitivity (.91) but poor specificity (.35). Similarly, a score of 10 resulted in good sensitivity (.89) but only slightly improved specificity (.47). Values were also calculated for a range of other cutoff scores (see Table 4). For this sample, a cutoff score of 15 appears to present the most balanced approach

Table 4. Sensitivity, Specificity, PPVs, and NPVs for CES-D-10 Cutoff Scores.

Cut score	Sensitivity	Specificity	PPV	NPV
8	0.91	0.35	0.79	0.61
9	0.90	0.40	0.80	0.60
10	0.89	0.47	0.82	0.62
11	0.87	0.51	0.83	0.58
12	0.84	0.56	0.84	0.56
13	0.81	0.64	0.86	0.55
14	0.79	0.70	0.88	0.55
15	0.76	0.75	0.89	0.53
16	0.74	0.79	0.90	0.52
17	0.68	0.83	0.92	0.49

Note. CES-D-10 = Center for the Epidemiological Studies of Depression Short Form (10 items); PPV = positive predictive value; NPV = negative predictive value.

to sensitivity (.76) and specificity (.75), with both values in the adequate range.

Discussion

To our knowledge, this is the first study to examine the psychometric properties of the CES-D-10 in a psychiatric sample. Overall, results suggest that the CES-D-10 is a reliable and valid measure for assessing symptoms of depression and evidences only adequate sensitivity and specificity in detecting a depression diagnosis in this acutely symptomatic psychiatric sample.

Results from the EFA suggested that a unidimensional factor provided the best fit to the data. The CFA specifying a single factor initially suggested a poor fit to the data; however, when we accounted for potential method effects by allowing the two reverse-scored items to covary, the resulting model fit was excellent. This is the first known study to account for the effect of reverse-wording on the CES-D-10. These results provide additional evidence that the previously identified "positive affect" factor may be better accounted for by method variance rather than a distinct factor. This addresses previous concerns about the construct validity of the "positive affect" factor (Cheng et al., 2006; Lee & Chokkanathan, 2008) and problems with factors consisting of only two items, given that factors with fewer than three items are generally weak and unstable (Costello & Osborne, 2005). Future work considering the influence of method factors on the factor structure of the CES-D-10 is needed, especially in psychiatric samples (see Rodebaugh et al., 2004, for an examination of method factors).

The CES-D-10 demonstrated strong convergent and divergent validity in this sample. Significant positive associations with depression and functioning and worry and negative associations with psychological well-being support our hypotheses. CES-D-10 scores were also significantly different between individuals with and without a diagnosis of a current major depressive episode, indicating sensitivity to depression even in a highly comorbid and acutely symptomatic sample. Overall, this suggests that the CES-D-10 may be a very effective instrument for assessing symptom severity in psychiatric samples.

The two cutoff scores of 8 and 10 derived from nonclinical samples functioned poorly in this sample. Both scores suffered from low rates of specificity and resulted in high rates of false positives. These results are likely due to the high levels of depression in the present sample, even among those without a current major depressive episode. To identify the optimal cutoff for our sample, we examined a wide range of potential cutoffs and found that a cutoff of 15 resulted in the most balanced combination of sensitivity (.76) and specificity (.75). Compared to previous work, estimates of sensitivity and specificity were lower in this sample compared to previous work, which estimated sensitivity at .91 and sensitivity at .92 (Zhang et al., 2012). This suggests that the CES-D-10 may have a substantial rate of false positives when used in highly symptomatic psychiatric samples. If the CES-D-10 is used as a screener in psychiatric samples, we recommend using a cutoff score of 16 over previously suggested scores of 8 or 10. Overall, the CES-D-10 is better suited for assessing levels of depressive symptoms and should be used with caution when used to screen for a depression diagnosis in psychiatric settings.

This study had several limitations. First, our examination of convergent validity may have been influenced by the use of the BASIS Depression and Functioning subscale and would have been strengthened by using a more widely studied measure of depression, such as the Beck Depression Inventory–II (Beck, Steer, & Brown, 1996), as well as by use of a non-self-report measure of depression. Second, although the sample was heterogeneous in its diagnostic presentation, it was relatively homogenous in terms of ethnicity. Examining the psychometrics of the CES-D-10 in a psychiatric sample with more ethnic and racial diversity would be beneficial in future work. Third, the ideal cutoff score for the CES-D-10 was calculated by giving equal weight to sensitivity and specificity. It is possible that a different cutoff would have been found, had sensitivity and specificity been given different weights. However, it is important to note that assigning equal weights is also standard in the literature and this facilitates comparison with previous studies. Finally, it should be noted that interrater reliability estimates for depression diagnoses were not available for the current study, although previous data from this population suggest adequate reliability, kappa = .69 (Kertz, Bigda-Peyton, Rosmarin, & Björgvinsson, 2012).

Despite these limitations, the present study extends the literature on the CES-D-10 by examining its psychometric properties in an acutely symptomatic, diagnostically heterogeneous psychiatric sample. We found that the CES-D-10

has overall strong psychometric properties. When using an alternate cutoff score of 15, the CES-D-10 functions adequately for screening for a clinical diagnosis of major depression in psychiatric settings, although other measures may perform better. In this sample, the CES-D-10 appeared to be a reliable and valid measure for assessing depression symptom severity in psychiatric settings.

Acknowledgment

Declaration of Conflicting Interests

Funding

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders, Fourth Edition*. Washington, DC: Author.

Andersen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *American Journal of Preventive Medicine, 10*, 77-84.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606. doi:10.1037/0033-2909.88.3.588

Blais, M. A., Lenderking, W. R., Baer, L., deLorell, A., Peets, K., Leahy, L., & Burns, C. (1999). Development and initial validation of a brief mental health outcome measure. *Journal of Personality Assessment, 73*, 359-373. doi:10.1207/S15327752JPA7303_5

Blais, M. R., Kehl-Fie, K., & Blias, M. A. (2008). The Schwartz Outcome Scale: A brief measure of psychological well-being. *Patient Reported Outcomes Newsletter, 39*, 19-21.

Bradley, K. L., Bagnell, A. L., & Brannen, C. L. (2010). Factorial validity of the Center for Epidemiological Studies Depression 10 in adolescents. *Issues in Mental Health Nursing, 31*, 408-412. doi:10.3109/01612840903484105

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.

Carpenter, J. S., Andrykowski, M. A., Wilson, J., Hall, L. A., Rayens, M. K., Sachks, B., & Cunningham, L. L. C. (1998). Psychometrics for two short forms of the Center for Epidemiologic Studies-Depression scale. *Issues in Mental Health Nursing, 19*, 481-494.

Cheng, S., & Chan, A. M. (2008). Detecting depression in Chinese adults with mild dementia: Findings with two versions of the Center for Epidemiologic Studies Depression Scale. *Psychiatry Research, 159*, 44-49.

Cheng, S., Chan, A. M., & Fung, H. H. (2006). Factorial structure of a short version of the Center for Epidemiologic Studies Depression Scale. *International Journal of Geriatric Psychiatry, 21*, 333-336. doi:10.1002/gps.1467

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indices for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

Cheung, Y., Liu, K., & Yip, P. F. (2007). Performance of the CES-D and its short forms in screening suicidality and hopelessness in the community. *Suicide and Life-Threatening Behavior, 37*, 79-88. doi:10.1521/suli.2007.37.1.79

Ciro, C., Ottenbacher, K., Graham, J., Fisher, S., Berges, I., & Ostir, G. (2012). Patterns and correlates of depression in hospitalized older adults. *Archives of Gerontology and Geriatrics, 54*, 202-205.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7), 1-9. Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Eisen, S. V., Normand, S., Belanger, A. J., Spiro, A., & Esch, D. (2004). The Revised Behavior and Symptom Identification Scale (BASIS-R): Reliability and validity. *Medical Care, 42*, 1230-1241. doi:10.1097/00005650-200412000-00010

Greenberg, P. E., Kessler, R. C., Birnbaum, H. G., Leong, S. A., Lowe, S. W., Berglund, P. A., & Corey-Lisle, P. K. (2003). The economic burden of depression in the United States: How did it change between 1990 and 2000? *Journal of Clinical Psychiatry, 64*, 1465-1475. doi:10.4088/JCP.v64n1211

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348-362. doi:10.1037/0022-3514.85.2.348

Helmreich, I., Wagner, S., Mergl, R., Allgaier, A., Hautzinger, M., Henkel, V., & Tadić, A. (2011). The Inventory of Depressive Symptomatology (IDS-C) is more sensitive to changes in depressive symptomatology than the Hamilton Depression Rating Scale (HAMD) in patients with mild major, minor or subsyndromal depression. *European Archives of Psychiatry and Clinical Neuroscience, 261*, 357-367. doi:10.1007/s00406-010-0175-1

Hopko, D. R., Reas, D. L., Beck, J., Stanley, M. A., Wetherell, J., Novy, D. M., & Averill, P. M. (2003). Assessing worry in older adults: Confirmatory factor analysis of the Penn State Worry Questionnaire and psychometric properties of an abbreviated model. *Psychological Assessment, 15*, 173-183. doi:10.1037/1040-3590.15.2.173

Howland, R. H. (1993). Chronic depression. *Hospital & Community Psychiatry, 44*, 633-639.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. doi:10.1080/10705519909540118

Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult. *Archives of Internal Medicine, 159*, 1701-1704.

Kertz, S. J., Bigda-Peyton, J., Rosmarin, D., & Björgvinsson, T. (2012). The importance of worry across diagnostic presentations: Prevalence, severity, and associated symptoms in a partial hospital setting. *Journal of Anxiety Disorders, 26*, 126-133.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., & Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association, 289*, 3095-3105. doi:10.1001/jama.289.23.3095

Kilbourne, A., Justice, A., Rabeneck, L., Rodriguez-Barradas, M., & Weissman, S. (2001). General medical and psychiatric comorbidity among HIV-infected veterans in the post HAART era. *Journal of Clinical Epidemiology, 54*(12), S22-S28.

Lee, A. Y., & Chokkanathan, S. (2008). Factor structure of the 10-item CES-D scale among community dwelling older adults in Singapore. *International Journal of Geriatric Psychiatry, 23*, 592-597. doi:10.1002/gps.1944

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*, 487-495. doi:10.1016/0005-7967(90)90135-6

Mitchell, A. J., Rao, S., & Vaze, A. (2011). Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *Journal of Affective Disorders, 130*(1-2), 26-36. doi:10.1016/j.jad.2010.07.028

Nishiyama, T., Ozaki, N., & Iwata, N. (2009). Practice-based depression screening for psychiatric outpatients: Feasibility comparisons of two types of Center for Epidemiologic Studies Depression Scales. *Psychiatry of Clinical Neuroscience, 63*, 632-638.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32, 396-402.

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

Richards, D. (2011). Prevalence and clinical course of depression: A review. *Clinical Psychology Review, 31,* 1117-1125.

Rodebaugh, T. L., Woods, C. M., Thissen, D. M., Heimberg, R. G., Chambless, D. L., & Rapee, R. M. (2004). More information from fewer questions: The factor structure and item properties of the original and brief Fear of Negative Evaluation Scale. *Psychological Assessment, 16*, 169-181.

Satorra, A., & Bentler P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Sharp, L. K., & Lipsky, M. S. (2002). Screening for depression across the lifespan: A review of measures for use in primary care settings. *American Family Physician, 66*, 1001-1009.

Sheehan, D. V., Lecrubier, Y., Sheehan, K., Amorim, P., Janavs, J., Weiller, E., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Suppl. 20), 22-33.

Weeks, J. W., Heimberg, R. G., Fresco, D. M., Hart, T. A., Turk, C. L., Schneier, F. R., & Liebowitz, M. R. (2005). Empirical validation and psychometric evaluation of the Brief Fear of Negative Evaluation Scale in patients with social anxiety disorder. *Psychological Assessment, 17*, 179-190. doi:10.1037/1040-3590.17.2.179

Zhang, W., O'Brien, N., Forrest, J. I., Salters, K. A., Patterson, T. L., Montaner, J. S., & Lima, V. D. (2012). Validating a shortened depression scale (10 Item CES-D) among HIV-positive people in British Columbia, Canada. *PLoS One 7*(7), e40793. doi:10.1371/journal.pone.0040793