Published in final edited form as:

Aging Ment Health. 2010 November ; 14(8): 962–970. doi:10.1080/13607863.2010.501060.

Confirmatory factor analysis of the Center for Epidemiologic Studies – Depression Scale in Black and White dementia caregivers

Crystal V. Flynn Longmire, PhD¹ and
Medical University of South Carolina

Bob G. Knight, PhD
University of Southern California

Abstract

Objectives—In order to better understand if measurement problems underlie the inconsistent findings that exist regarding differences in depression levels between Black and White caregivers, this study examined the factor structure and invariance of the Center for Epidemiologic Studies-Depression scale (CES-D).

Method—A confirmatory factor analysis of the 20-item CES-D was performed on a sample of 167 Black and 214 White family caregivers of older adults with dementia from Los Angeles County.

Results—The relationships between the 20 items and the four factors, as well as the relationships among each of the factors, were equivalent across both caregiver groups, indicating that the four-factor model fit the data for both racial groups.

Conclusion—These findings offer further evidence that the standard four-factor model is the best fitting model for the CES-D and is invariant across racial groups.

Keywords

depressive symptoms; dementia caregivers; structural equation model; factorial invariance

Introduction

The responsibility of ongoing care for family members with dementia can have serious negative effects on the caregivers (Schulz, O'Brien, Bookwala, & Fleissner, 1995). Dementia caregivers provide significantly more hours of care per week than nondementia caregivers, and they report more employment complications, caregiver strain, mental and physical health problems, and family conflict, while having less leisure and family time (Ory, Hoffman III, Yee, Tennstedt & Schulz, 1999). In a review of the dementia caregiver literature, the large majority of studies showed elevated levels of depressive symptomatology; further, studies including diagnostic interviews found high rates of clinical depression (Schulz et al., 1995).

Whether the effects of the stress of caregiving on mental health vary among ethnic groups has been a key question in the caregiving literature and racial and ethnic differences in outcomes are a common finding (see reviews by Dilworth-Anderson, Canty Williams, &

¹Corresponding author: Crystal V. Flynn Longmire - Mailing address: 5900 Core Road, Suite 203, N. Charleston, SC 29406, USA; flynnlo@musc.edu; Phone: 843-740-1592 Fax: 843-740-6113.

Gibson, 2002; Janevic & Connell, 2001). In particular, some researchers have found Black caregivers report lower mean levels of depression (Clay, Roth, Wadley, & Haley, 2008; Sleath, Thorpe, Landerman, Doyle, & Clipp, 2005; Sorensen & Pinquart, 2005), while other researchers have found that Black and White caregivers have similar mean levels of depressive symptoms (Hinrichsen & Ramirez, 1992; Knight, Flynn Longmire, Dave', Kim, & David, 2007; Knight & McCallum, 1998). Overall, Black caregivers appear to have lower or equal, but not higher, levels of depression than White caregivers.

Common history, normative expectations, and intergenerational transmission of shared values distinguish ethnic groups, and factors posited to contribute to findings that vary by ethnicity include social class, racial discrimination and genetics (Pinquart and Sorensen, 2005). Another plausible explanation is that the measures do not operate the same way across racial/ethnic groups and that measurement studies are needed to tease apart the contribution of measurement variance versus the variance due to racial/ethnic/cultural differences (Janevic and Connell, 2001). One method for studying measurement is to examine a scale's factor structure and to test if the structure fits and the parameters are equal for two or more groups. If the measurement model or structural model is equivalent across the groups of interest (e.g., gender, ages, ability; Byrne, 2001), this is called multigroup invariance.

Testing an instrument's factorial structure for multigroup invariance is necessary in order to trust that differences found between groups on a particular variable are not due to the instrument measuring differing sets of concepts depending on the group responding. There are a variety of sources of variance in any analysis, so the intent of invariance testing is to eliminate measurement inequivalence, or group differences in the relationship between test scores and the measure's underlying factor structure (Lubke, Dolan, Kelderman, & Mellenbergh, 2003), as a significant source of variance in the statistical analyses. If the measurement structure varies across groups, it raises the question of whether the concepts initially hypothesized to be measured by an instrument are in fact being measured in the group(s) on which it was not initially developed.

In comparing ethnic groups in caregiver research, one of the gaps in knowledge that must be addressed is whether or not the commonly used measures are appropriate for diverse groups. The Center for Epidemiologic Studies-Depression Scale (CES-D; Radloff, 1977) is one of the most extensively used measures of depressive symptomatology in caregiver research. Moreover, the CES-D has consistently been employed to compare Black and White caregivers on levels of depression. However, while many studies include mean differences in CES-D scores and relationships between depression and caregiver characteristics, less research has assessed the factor structure of the scale across groups. Hence, the present study focuses on the psychometric properties of the CES-D and whether this scale operates similarly in Black and White dementia caregivers by examining the invariance of the CES-D factor structure for the two groups.

Among the studies of the CES-D that include factor analyses, most support a standard four-factor structure, but there are a few studies that support other models. There have been even fewer studies that specifically sample dementia caregivers, a population with elevated levels of depressive symptoms. Only one of these studies considered racial comparisons; however, there have been non-caregiver studies that have included multiple racial, cultural, and ethnic groups as well as studies with older adult samples. The following are examples of studies that investigate the factor structure of the CES-D, starting with its development.

Previous Analysis of the Center for Epidemiologic Studies-Depression Scale

The first factor analysis of the CES-D was done during the development of the CES-D. It was an exploratory factor analysis to consider what concepts (factors) were represented by the scale's items and what the structure of those factors might be (Radloff, 1977). The original four factors found were called Depressed Affect, Positive Affect, Somatic and Retarded Activity, and Interpersonal Feelings. Initially, only sixteen of the scale's twenty items were used in the four factors. Radloff mentioned that the high internal consistency of the scale argued against undue emphasis on separate factors. However, Radloff and Teri (1986) stated that though factor analysis for independent factors should be interpreted with caution for scales with very high internal consistency, factor analysis does give some indication about the clustering of subsets of items. Further, they concluded that, for some purposes, it can be useful to score subscales based on the four factors. In our study, the standard four-factor structure refers to Radloff's four factors, and, as seen in later factor analysis studies, including all 20 items.

The standard four factors were validated in two samples of adults (Hertzog, Van Alstine, Usala, Hultsch, & Dixon, 1990). A sample from the U.S. included younger and older adults ages 20–80, and a Canadian sample included older adults ages 55 to 78. Using confirmatory factor analysis, they found the four-factor model fit both samples. Hertzog et al. also analyzed the factor pattern for invariance between three groups, U.S. younger, U.S. older and Canadian older adults, and found invariant unstandardized item factor loadings across samples and age groups. Their results supported the measurement validity of the CES-D for use as a screening tool for depression in older adults, and the use of its four subscales. The standard structure was also confirmed in a sample of frail, older adults (Davidson, Feldman, and Crawford, 1994) for whom the scale was found to be highly reliable with a Cronbach's $\alpha = .86$, and with good model fit.

Another study extended the examination of the factor structure of the CES-D to five Asian countries (Mackinnon, McCallum, Andrews and Anderson, 1998). The study considered cultural differences in reporting of depressive symptoms among older people in Indonesia, Korea, Myanmar, Sri Lanka, and Thailand by analyzing seventeen of the twenty CES-D items (three items were viewed as culturally inappropriate). In each of the five countries, the four-factor model fit better than a model with one general factor. Mackinnon et al. concluded that the factor structure for the Asian countries was similar to the findings in Western settings and that the CES-D has comparable measurement properties across countries.

Several studies have also confirmed model fit for the CES-D four-factor model in African American samples with varying levels of invariance. Hales et al. (2006) validated the four-factor structure and found invariant factor loadings in samples of Black and White adolescent girls. Nguyen, Kitner-Triolo, Evans and Zonderman (2004) examined three samples of African Americans where socioeconomic status varied, and found that the four-factor model fit, and that the number of factors and the pattern of factor loadings were equivalent across all three groups. However, when equality constraints were placed on the factor loadings, the magnitudes of the loadings were not equivalent across groups. Lastly, a comparison of African American women based on age (below 60 years /60 years and above) showed the four-factor solution to be a good fit, but when factor loadings were constrained to be equal, the magnitude of the item loading onto the factors were not equivalent across the age groups (Williams et al., 2006).

There have recently been confirmatory factor analysis studies with caregiver samples (O'Rourke, 2003; Roth, Ackerman, Okonkwo & Burgio, 2008). O'Rourke examined the invariance of the CES-D four-factor model in French vs. English language versions among

Canadian dementia caregivers (2003) and replicated the four-factor model used by Hertzog et al. (1990). He found that the factor structure was similar across language groups, and concluded that his findings supported the use of the CES-D in cross-cultural and comparative research with English- and French-speaking adults.

Roth et al. (2008) reported that out of the analyses of one-, two-, three- and four-factor models, the four-factor model fit best among White, African American, and Hispanic dementia caregivers in the REACH data set despite significant differences in mean CES-D scores among the three groups. They also found that three of the factor loadings and one factor correlation were not invariant across the race groups. However, they explained that this often happens with large samples and the pattern of their results is called “approximate invariance,” such that the group differences are minimal and unlikely to cause significant biases when the instrument is used in regular practice.

Problem Statement and Research Questions

In spite of several confirmatory studies of the standard structure, the structure of the CES-D has not been extensively studied in African Americans in general and even less so in African American caregivers. These studies have found different factor structure solutions for the CES-D for different ethnic groups, or lack of invariance across groups. Given the lack of research on the invariance of the CES-D in African American caregivers and the variability in comparisons with mean CES-D scores between Blacks and Whites, examining the factorial invariance of the CES-D is important to research on African American caregiving.

Consequently, the first goal of the present paper was to examine the factor structure of the CES-D in Black and White dementia caregivers to test our hypothesis that the pattern, i.e., the number and configuration, of the factors for the CES-D are the same across groups. The standard four-factor structure of the CES-D was analyzed without any restrictions on the parameters. The second goal of this study was to determine whether measurement qualities of the CES-D were similar in Blacks and Whites, or whether measurement inconsistency underlies the discrepant findings regarding depression in Black and White dementia caregivers. Therefore, beyond fit of the overall pattern of the CES-D four-factor model, the magnitudes of the item loadings on the factors and then the factor covariances were tested for equivalence between Black and White caregivers. The assumption was that the model would still fit both of the two caregiver groups, even when the factor loading pattern (hypothesis 2) and the factor covariances (hypothesis 3) were set as equal. The endeavor here was not to find a new factor structure, but to confirm whether or not the standard four-factor structure fits and operates similarly for caregivers from a minority group as well as caregivers from a non-minority group.

Methods

Participants

This study was an analysis of secondary data. The participants were Black and White dementia caregivers obtained by combining data from two samples. First, a convenience sample was collected from 1990–1996 through the Research Training and Information Transfer Core of the Alzheimer’s Disease Research Center – Southern California consortium at the University of Southern California (USC). The sample was recruited from a variety of referral sources such as the Alzheimer’s Association of Los Angeles County and the Los Angeles Caregiver Resource Center. From 1990–1993 participants (n=159) qualified if they were the primary caregivers of the older adult with dementia and if they were adults themselves (at least 18 years of age). Beginning in 1994, enrollment criteria changed and

participants (n=95) had to be at least 50 years old, and either living with or providing at least 8 hours of care per week for the care recipient who suffered from chronic memory loss.

Second, a probability sample was obtained from the Stress, Ethnicity and Caregiving study (SECS), which was also conducted at USC (n=160). Data were obtained from 2001–2003. This random sample of caregivers was obtained through phone calls made to homes in Los Angeles County. The households called were located through census tracts which had high percentages of older adults residing in them (10% or more 65 years or older), and were predominately (60% or higher) African American or White, non-Hispanic. To decrease possible socioeconomic confounds, the tracts selected also had to have an average household income below the median for Los Angeles County. Caregivers were included as participants in SECS if they said they were currently caring for a family member (or someone they considered family) who was over the age of 55 and had abnormal memory problems or was diagnosed with dementia.

In the combined data set, all of the participants were primary, not secondary, caregivers. The primary caregiver is the person most responsible for the older adult with dementia or severe memory problems. All participants were asked the same questions in the same format. Caregivers chose the interview times. The interviews were done either in the home of the caregiver or at the University, depending on the caregiver's preference. The vast majority (> 90%) of the interviews were conducted at home in both the convenience sample and the SECS sample. Home interviews are considered a way to decrease bias toward less distressed caregivers (Knight et al., 2000). The samples of caregivers were combined to increase power for the analysis; such combination was logical since all three were the same type of caregivers (dementia), lived in the same geographical area (Los Angeles area), and were tested using the same measure (CES-D) administered by the same group of people (USC research team).

Measures

Sociodemographic characteristics—Age and gender were included in the analyses. Other variables (e.g., income and health) were not consistently asked across the surveys, and thus were not included.

Depression—The Center for Epidemiologic Studies Depression Scale (Radloff, 1977) is a 20-item self-report scale developed to screen for depressive symptomatology in the general population. Each response was scored from 0 to 3. Four items are worded in a positive direction. These four items are reverse coded before summing the items for the total score. Total scores range from 0 to 60 with higher scores indicating more depressive symptomatology. Scores above a standard cutoff of 16 are said to be indicative of clinical levels of depression. Internal consistency, for all 20 items, was .88 for the Black sample, and .85 for the White sample.

Missing Data

Of the 414 caregivers interviewed, 231 White caregivers and 183 Black caregivers had CES-D data. All cases with missing data were deleted. In total, 16 cases (8.7%) were deleted from the Black CES-D file (n=167), and 17 cases (7.4%) were deleted from the White CES-D file (n=214). Casewise deletion was used since less than 10% of the sample size had missing data. Selection bias was not considered an issue as there was no systematic pattern to the missing data.

Analysis

Confirmatory factor analysis using AMOS 4.0 was used. The CES-D analyses examined the standard 20-item, 4-factor model. For the first hypothesis, to test that the pattern of items and factors for this scale was equivalent across groups, both the Black and the White caregiver samples were analyzed simultaneously for model adequacy with no parameter constraints. The second and third hypotheses related to examining the invariance of the measurement qualities of the CES-D. To test the second hypotheses, again both caregiver samples were analyzed simultaneously as in the analysis for the first hypothesis, but this time, however, the factor loadings were constrained to be equal across the two groups. The test for the third hypothesis had both factor loadings and the factor covariances among the four factors constrained equally across the two groups. These analyses focus on metric equivalence (Posner, Stewart, Marín, & Pérez-Stable, 2001). Thus, the equality of the item error terms was not tested. Except in particular instances, when for example the invariant reliability of a scale is of interest, the equality of error variances and covariances is considered an overly restrictive test of the data (Byrne, 2001). The estimation procedure employed in these analyses was maximum likelihood (ML).

Our main interest in the confirmatory factor analysis was the degree to which the hypothesized measurement model adequately described, or fit, the sample data. In general, indices of model adequacy fall into the categories of model fit, model comparison, or model parsimony (Schumacker and Lomax, 1996). We include each type to compare and assess the fit of the CES-D four-factor model to our samples of dementia caregivers.

The chi-square (χ^2) goodness-of-fit statistic is a test of overall model fit. However, sensitivity of the χ^2 to sample size and departures from multivariate normality have led to the use of the χ^2/df (degrees of freedom) ratio. A small value indicates good fit, and a large one indicates poor fit (Byrne, 2001). Some researchers allow values for χ^2/df as large as 5 as being an adequate fit, but conservative use calls for rejecting models with relative chi-square greater than 2 or 3 (Garson, 2004). In these analyses, the more conservative standard was used.

The parsimony goodness-of-fit index (PGFI) considers the complexity, or number of estimated parameters, of the hypothesized model while also assessing the relative amount of variance and covariance in the sample that is explained by the model. The PGFI tends to have lower values than the generally accepted standard levels for other indices of fit (i.e., .90), and values in the .50s are acceptable (Byrne, 2001).

The comparative fit index (CFI) measures the comparative reduction in lack of fit by a hypothesized model versus a baseline model (Hoyle & Panter, 1995). Originally a value $>.90$ was advised as a cutoff for a well-fitting model, though more recently a value close to .95 has been proposed recommended (Hu & Bentler, 1999). This index is independent of sample size and takes model complexity into account.

The root mean square error of approximation (RMSEA), an index of model parsimony, measures the lack of fit per degree of freedom (MacCallum, 1995). Values less than .05 indicate good fit. Values up to .08 indicate reasonable errors of approximation in the population, while those above .1 represent poor fit. With some agreement on cutpoints, statisticians also provide a cautionary note that when sample size is small, the RMSEA has a tendency to overreject true population models (Byrne, 2001). However, a 90% confidence interval (CI) around RMSEA assists with deciding the precision of the RMSEA, where a narrow confidence interval argues for a more precise RMSEA. As well, when the RMSEA is less than .05 there is an associated test of closeness of fit. For this test of close fit, probability values should be greater than .05.

Along with the fit indices, a chi-square difference test was used to consider whether or not constrained parameters should be considered invariant. The difference in chi-square for competing or nested models is itself chi-square distributed with degrees of freedom equal to the consequent difference in degrees of freedom. If the difference is found not to be significant then the hypothesis of an invariant pattern of loadings is considered tenable (Byrne, Shavelson & Muthén, 1989).

Results

Characteristics of the Sample

White caregivers were significantly older than Black caregivers, the majority of both groups of caregivers were female, and the two groups did not differ significantly on their CES-D total scores. Both groups had mean scores slightly above the CES-D threshold of 16 for clinical depression. This coincides with the literature that has found dementia caregivers to have high levels of depressive symptomatology. See Table 1 for specific results. Skewness is typical in non-patient samples, as there is generally a larger proportion of low scores (Radloff, 1977). The data were skewed for both groups, but the distributions did not depart substantially from normality (i.e., skewness was < 2). The factor analysis assumption of normality was not violated.

Factor Analysis Results

The standard four factors were used in these analyses: Depressed Affect, Somatic and Retarded Activity, Positive Affect, and Interpersonal Feelings. Table 2 shows the items associated with each factor and their loadings on the factor, along with error terms for factors and items.

Hypothesis 1—For the test of the first hypothesis that the CES-D would have four factors in both groups, the χ^2/df was less than two, indicating good fit. The PGFI was .671, and the CFI was .941. The PGFI provided support for model fit, and the CFI showed an improvement in fit using this model vs. an independence model. The RMSEA was .048 and fell within a 90% confidence interval that ranged from .042 to .054, which represented a good degree of precision. The probability value associated with a test of close fit for the RMSEA in the population was .731. The number of factors and their associated items for this CES-D model was equivalent across racial groups.

Hypothesis 2—The test of the second hypothesis regarding equivalent factor loadings yielded a χ^2 of 630.128 with 344 degrees of freedom ($p < .001$). The degrees of freedom increased by 16 when the item regression weights were constrained to be equal across groups (16 CES-D items constrained equal and 4 still fixed at 1 to set factor scales). The χ^2/df was also less than two for this analysis, an indicator of good fit. The PGFI was .701 and CFI was .941. Both the PGFI and the CFI upheld the assumption that the model fit the data. The RMSEA remained almost the same as in the previous analysis at .047. The CI for the RMSEA was .041 to .053, and the test of close fit had a p-value of .831. The indices taken together suggested that the model fit the data. The standard CES-D model seemed equivalent in terms of factor composition and loadings across groups.

Hypothesis 3—For the third hypothesis, the parameters for both samples were all significant and in the right direction, and the degrees of freedom increased by 6 when the factor covariances were constrained equally across groups. The χ^2 was 636.046 for 350 degrees of freedom with a p-value $< .001$, and the χ^2/df suggested satisfactory fit. The correlation between Depressed Affect and Somatic and Retarded Activity (.67) was the highest among the factor covariances for both the Black and the White caregivers, and the

correlation between Somatic and Retarded Activity and Interpersonal Feelings (.37) was the weakest. The PGFI was .712, and the CFI remained the same at .941. The two indices indicated model fit. The RMSEA was good at .046. The CI ranged from .041 to .052, and the test of close fit had a $p = .845$. Altogether, the indices were evidence for model fit. The relationships between the four factors seemed to be equivalent for the CES-D model. Invariance was shown for this model. Tables 2 – 4 show the results. Table 5 offers a comparison of the difference in chi-square as the amount of restriction (constraint) within the model increased.

Post hoc analysis

After testing the hypotheses, a more stringent analysis of the CES-D four-factor structure was done. In this analysis, we constrained the factor variances to test whether there was invariance at this high level of constraint. As in the other three CES-D model analyses, the estimation process was completed, and all parameters were significant and in the correct direction (Tables 4 and 5). The chi-square was 643.785 with 354 df and $p < .001$. The PGFI was .719 and the CFI was .945. Both indices showed model fit. The RMSEA was .046, and represented good fit. With even stronger constraints, this model fit the data. The difference in chi-square for this model compared to the baseline model was not significant, so the hypothesis that factor loadings, covariances and variances were invariant was considered tenable.

Summary of findings

In general, the standard model fit in all four analyses. The χ^2/dfs were less than two. The PGFIs were above .50, and the CFIs were consistently greater than .90 and close to .95. The RMSEA values consistently indicated good fit ($< .05$). They had good degrees of precision and tests of close fit. All models had complete estimation.

Analyses for hypotheses 2 and 3 showed improved fit for the CES-D four-factor model when constraints were added. Though the χ^2 did increase in these analyses, the increases in chi-square were not significant as compared to the unconstrained baseline analysis. Further, the post hoc analysis showed that at a level of equality constraints beyond what was needed to answer the research questions, the model still fit both groups. There was evidence of metric invariance regarding the number and composition of factors, the factor loadings, the covariance between the factors, and the factor variances for the CES-D.

Discussion

A significant part of understanding the caregiver experience is to ensure that the scales used are reliably measuring constructs of interest. Given the wide use of the CES-D in caregiver literature and greater prevalence of Black and White caregiver comparisons, this study focused on the metric invariance of the CES-D in Black and White dementia caregivers. There were two main goals (1) to see if the standard CES-D four-factor structure held in a sample of Black caregivers as well as White caregivers, and (2) to see if the CES-D measured the same constructs equivalently in both samples. The CES-D four-factor structure fit the data, and was invariant between groups of Black and White dementia caregivers.

These results confirmed the standard four-factor structure and supported similar findings. The standard factor structure has been confirmed in middle-aged and older adults (Hertzog et al., 1990), frail elders (Davidson et al., 1994), Asian adults ages 60 and over (Mackinnon et al., 1998), African Americans of various ages (Hales et al., 2006; Nguyen et al., 2004; Williams et al., 2006), Canadian dementia caregivers (O'Rourke 2003, 2005) and samples of Black, White and Hispanic dementia caregivers (Roth et al., 2008), but at various levels of

equivalence, if at all. This study expanded upon the information about the psychometric properties of the CES-D by examining Black and White groups among dementia caregivers in a U.S. sample in which the CES-D scores were equal between blacks and Whites and finding measurement invariance for the four CES-D factors, the factor loadings, the factor covariances and the factor variances.

This study adds to the invariance literature by providing results of an examination of metric invariance for the CES-D with Black and White dementia caregivers at higher levels than in previous similar studies. The CES-D had similar psychometric properties and operated equivalently for both Black and White caregivers. In particular, these findings occurred in a sample where varying recruitment techniques were employed, which has been discussed as a possible factor when factor differences were found between these groups (Roth et al., 2008). This study provides evidence that findings regarding the CES-D between Black and White caregivers are not likely due to inequivalence. This is a significant point for the majority of the comparative literature on caregivers which includes Black and White comparisons.

Limitations

It is necessary to address several limitations of this study. First, for the Black sample of caregivers, the sample size was less than 200. Though some fit indices, such as the chi-square, are more likely to reject models when the sample size is large, they are also lenient when the sample size is small, possibly making models easier to fit. We tried to address this issue by reporting a variety of fit indices; however, this is a limitation that must be considered when reviewing these findings. Larger samples would also allow for analysis with more variables or subgroups, which could aid investigations of what underlies group differences outside of measurement concerns. The study sample was regional to the western U.S., with most participants living in the Los Angeles County area. Almost forty percent of the total sample was from areas whose average income was below the County mean, and though not everyone in the sample had lower incomes, this may have had an effect on this caregiver sample. As such, these results may not be generalizable to the entire population of dementia caregivers.

There is variation regarding methods required to demonstrate invariance, and some researchers would call the invariance found in this study “weak factorial invariance,” as opposed to “strong or strict factorial invariance,” since the specific factor means and the residual variances were not included among the constraints tested (Meredith & Teresi, 2006). On the other hand, since these findings are mainly to be applied to scientific research of individual behavior, it may be that pattern invariance is fundamental while strong or strict invariance are less important (Meredith & Teresi, 2006). Further, the methods in this study similar to those employed elsewhere in the literature (Nguyen et al., 2004; Roth et al., 2008; Williams et al., 2006), and given the relative dearth of information regarding factor structures of scales, it would seem that studies of invariance even at the less strict levels can add to the body of knowledge available.

Lastly, these analyses were conducted on the twenty-item version of the CES-D. While promising for other versions, these findings cannot be extrapolated to other versions with fewer or different items, as these issues directly affect factor structure. Each version would need to be analyzed to first determine their factor structure and then to uncover if that structure was invariant across groups.

Implications

Reliable and culturally valid screening instruments are vital to clinical practice (Mui, Burnette, & Chen, 2001). These findings assist service providers and practitioners in that

they do not have to use different scales to screen Black and White caregivers. The findings do not mean that Black and White caregivers have the same needs, but they do show that the CES-D is useful for measuring depressive symptoms in both groups of caregivers. In addition, in order to know prevalence rates and to devise treatment plans and services for caregivers, it is important that scales provide accurate detection. Erroneous conclusions drawn from measures with dissimilar properties across groups can also have wide-ranging implications for public health policy (Posner et al., 2001).

References

- Byrne, BM. Structural Equation Modeling with AMOS: Basic concepts, applications, and programming. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
- Byrne BM, Shavelson RJ, Muthén B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*. 1989; 105:456–466.
- Clay OJ, Roth DL, Wadley VG, Haley WE. Changes in social support and their impact on psychosocial outcome over a 5-year period for African American and White dementia caregivers. *International Journal of Geriatric Psychiatry*. 2008; 23:857–862. [PubMed: 18338341]
- Davidson H, Feldman PH, Crawford S. Measuring depressive symptoms in the frail elderly. *Journals of Gerontology*. 1994; 49:159–164.
- Dilworth-Anderson P, Canty Williams I, Gibson BE. Issues of Race, Ethnicity, and Culture in Caregiving Research: A 20-Year Review (1980–2000). *The Gerontologist*. 2002; 42:237–272. [PubMed: 11914467]
- Garson, GD. Structural Equation Modeling Example Using WinAMOS (webpage). 2004. <http://www2.chass.ncsu.edu/garson/pa765/semAMOS1.htm>
- Hales DP, Dishman RK, Motl RW, Addy CL, Pfeiffer KA, Pate RR. Factorial validity and invariance of the center for epidemiologic studies depression (CES-D) scale in a sample of Black and White adolescent girls. *Ethnicity and Disease*. 2006; 16:1–8. [PubMed: 16599341]
- Hertzog C, Van Alstine J, Usala PD, Hultsch DF, Dixon R. Measurement properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations. *Psychological Assessment*. 1990; 2:64–72.
- Hinrichsen GA, Ramirez M. Black and white dementia caregivers: A comparison of their adaptation, adjustment, and service utilization. *The Gerontologist*. 1992; 32:375–381. [PubMed: 1500003]
- Hoyle, RH. The structural equation modeling approach: Basic concepts and fundamental issues. In: Hoyle, RH., editor. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage; 1995. p. 158–176.
- Hoyle, RH.; Panter, AT. Writing about structural equation models. In: Hoyle, RH., editor. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage; 1995. p. 158–176.
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A multidisciplinary Journal*. 1995; 6:1–55.
- Janevic MR, Connell CM. Racial, Ethnic, and Cultural Differences in the Dementia Caregiving Experience. *The Gerontologist*. 2001; 41:334–347. [PubMed: 11405431]
- Knight BG, Flynn Longmire CV, Dave J, Kim JH, David S. Mental health and physical health of family caregivers for persons with dementia: A comparison of African American and white caregivers. *Aging & Mental Health*. 2007; 11:538–546. [PubMed: 17882592]
- Knight BG, McCallum TJ. Heart rate reactivity and depression in African-American and white dementia caregivers: Reporting bias or positive coping? *Aging & Mental Health*. 1998; 2:212–221.
- Knight BG, Silverstein M, McCallum TJ, Fox LS. A socio-cultural stress and coping model for mental health outcomes among African American caregivers in Southern California. *Journal of Gerontology: Psychological Sciences*. 2000; 55B:142–150.

- Lubke GH, Loan CV, Kelderman H, Mellenbergh GJ. Weak measurement invariance with respect to unmeasured variables: An implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology*. 2003; 56:231–248. [PubMed: 14633334]
- MacCallum, RC. Procedures, strategies, and related issues. In: Hoyle, RH., editor. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage; 1995. p. 16-36.
- Mackinnon A, McCallum J, Andrews G, Anderson I. The Center for Epidemiological Studies Depression Scale in older community samples in Indonesia, North Korea, Myanmar, Sri Lanka, and Thailand. *Journal of Gerontology: Psychological Sciences*. 1998; 53B:343–352.
- Meredith W, Teresi JA. An essay on measurement and factorial invariance. *Medical Care*. 2006; 44:S69–S77. [PubMed: 17060838]
- Mui AC, Burnette D, Chen LM. Cross-cultural assessment of geriatric depression: A review of the CES-D and the GDS. *Journal of Mental Health and Aging*. 2001; 7:137–164.
- Nguyen HT, Kitner-Triolo M, Evans MK, Zonderman AB. Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. *Psychiatry Research*. 2004; 126:177–187. [PubMed: 15123397]
- O'Rourke N. Research Note: Equivalence of French and English Language Versions of the Center for Epidemiologic Studies-Depression Scale (CES-D) among Caregivers of Persons with Dementia. *Canadian Journal on Aging*. 2003; 22:323–329.
- O'Rourke N. Factor Structure of the Center for Epidemiologic Studies--Depression Scale (CES--D) Among Older Men and Women Who Provide Care to Persons with Dementia. *International Journal of Testing*. 2005; 5:265–277.
- Ory MG, Hoffman RR III, Yee JL, Tennstedt S, Schulz R. Prevalence and impact of caregiving: a detailed comparison between dementia and nondementia caregivers. *The Gerontologist*. 1999; 39:177–185. [PubMed: 10224714]
- Pinquart M, Sorensen S. Ethnic differences in stressors, resources, and psychological outcomes of family caregiving: A meta-analysis. *The Gerontologist*. 2005; 45:90–106. [PubMed: 15695420]
- Posner SF, Stewart AL, Marín G, Pérez-Stable EJ. Factor validity of the Center for Epidemiological Studies Depression Scale (CES-D) Among Urban Latinos. *Ethnicity and Health*. 2001; 6:137–144. [PubMed: 11488294]
- Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1:385–401.
- Radloff LS, Teri L. Use of the Center for Epidemiological Studies-Depression Scale with older adults. *Clinical Gerontologist*. 1986; 5:119–136.
- Roth DL, Ackerman ML, Okonkwo OC, Burgio LD. The four-factor model of depressive symptoms in dementia caregivers: A structural equation model of ethnic differences. *Psychology and Aging*. 2008; 23:567–576. [PubMed: 18808246]
- Schulz R, O'Brien AT, Bookwala J, Fleissner K. Psychiatric and physical morbidity effects of dementia caregiving: Prevalence, correlates, and causes. *The Gerontologist*. 1995; 35:771–791. [PubMed: 8557205]
- Schumacker, RE.; Lomax, RG. *A Beginner's Guide to Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates; 1996.
- Sleath B, Thorpe J, Landerman LR, Doyle M, Clipp E. African-American and White Caregivers of Older Adults with Dementia: Differences in Depressive Symptomatology and Psychotropic Drug Use. *Journal of the American Geriatrics Society*. 2005; 53:397–404. [PubMed: 15743280]
- Sörensen S, Pinquart M. Racial and ethnic differences in the relationship of caregiving stressors, resources, and sociodemographic variables to caregiver depression and perceived physical health. *Aging & Mental Health*. 2005; 9:482–495. [PubMed: 16024408]
- Tran TV. Exploring the equivalence of factor structure in a measure of depression between black and white women: Measurement issues in comparative research. *Research on Social Work Practice*. 1997; 7:500–517.
- Williams CD, Taylor TR, Makambi K, Harrell J, Palmer JR, Rosenberg L, Adams-Campbell LL. CES-D four-factor structure is confirmed, but not invariant, in a large cohort of African American women. *Psychiatry Research*. 2007; 150:173–180. [PubMed: 17291596]

Table 1

Demographic variables of Caregivers and Dependent Variables by race.

	White (n=225)		Black (n=175)		
Background characteristics	SD		SD		Chi-Square T-test
Female Caregivers	68.9%	--	76.0%	--	2.47
Age ^a	62.88	(13.57)	57.48	(14.44)	3.83 ***
Self-reported depression-CESD	17.74	(9.64)	16.06	(11.31)	1.57

Note: Categorical variable Female is reported with percentage and a chi-square test. Other entries reported as means and standard deviations in parentheses with t-tests.

^aWhite n=224, Black n=173 for age only.

p<.001

Table 2

Unstandardized factor loadings and factor covariances for the CES-D constrained equal across Black and White caregiver groups

CES-D-4 Factor Names and 20 Items		Factor Loadings	Factor/ Item Error Terms
<u>Factor:Depressed Affect</u>			.91
Item #	Item Name		
3	Could not shake the blues	1.00	.39
6	Felt Depressed	1.06	.44
9	Thought my life had been a failure	.67	.39
10	Felt fearful	.83	.35
14	Felt lonely	1.00	.41
17	Had crying spells	.83	.60
18	Felt sad	1.06	.36
<u>Factor:Somatic and Retarded Activity</u>			.62
1	Bothered by things that don't usually bother me	1.00	.55
2	Appetite was poor	.80	.59
5	Trouble keeping my mind on tasks	.95	.51
7	Everything I did was an effort	1.16	.52
11	Sleep was restless	.98	1.00
13	Talked less than usual	1.00	.60
20	Could not get "going"	1.10	.42
<u>Factor:Positive Affect</u>			1.35
4	Good as other people	1.00	1.41
8	Hopeful about the future	.76	.90
12	Was happy	.95	.41
16	Enjoyed life	1.08	.45
<u>Factor:Interpersonal Feelings</u>			.32
15	People were unfriendly	1.00	.22
19	Felt that people disliked me	1.04	.17

Note: Only one set of coefficients reported since parameters for Black group and White group are equal.

Table 3

Factor covariances for CES-D when factor loadings and covariances are constrained equal between Black and White caregivers

	Depressed Affect	Somatic and Retarded Activity	Positive Affect	Interpersonal Feelings
Depressed Affect	--	.67	.65	.44
Somatic and Retarded Activity	--	--	.54	.37
Positive Affect	--	--	--	.42
Interpersonal Feelings	--	--	--	--

Table 4

Fit statistics for confirmatory factor analyses of the CES-D four factors with various levels of constraint including both samples.

Model	χ^2	df	Ratio(Chi/df)	PGFI	CFI	RMSEA
Model 1a	611.377	328	1.863954	0.671	0.941	0.048
Model 2b	630.128	344	1.831767	0.701	0.941	0.047
Model 3c	636.046	350	1.817274	0.712	0.941	0.046
Model 4d+	643.785	354	1.818602	0.719	0.945	0.046

Note: N = 381. χ^2 = Chi square; df= degrees of freedom; PGFI= parsimony goodness-of-fit; CFI = comparative fit index; RMSEA = root mean square error of approximation.

a = no constraints; b = factor loadings (item regression weights) constrained equal; c = factor loadings and factor covariances constrained; d=factor loadings, covariances and variances constrained; + = post hoc analysis

Table 5

Chi-square difference tests for the CES-D four-factor models at various levels of constraint.

Model	χ^2	df	p-value
Model 1a			Baseline model
Model 2b	18.75	16	>0.05
Model 3c	24.67	22	>0.05
Model 4d+	32.41	26	>0.05

Note: χ^2 = Chi square; df= degrees of freedom; a = no constraints; b = factor loadings (item regression weights) constrained equal; c = factor loadings and factor covariances constrained; d=factor loadings, covariances and variances constrained; + = post hoc analysis