Published in final edited form as:

Psychol Aging. 2008 September ; 23(3): 567–576. doi:10.1037/a0013287.

The Four-Factor Model of Depressive Symptoms in Dementia Caregivers: A Structural Equation Model of Ethnic Differences

David L. Roth,

University of Alabama at Birmingham

Michelle L. Ackerman,

University of Alabama at Birmingham

Ozioma C. Okonkwo, and

University of Alabama at Birmingham

Louis D. Burgio

University of Alabama

Abstract

Previous studies have suggested that 4 latent constructs (depressed affect, well-being, interpersonal problems, somatic symptoms) underlie the item responses on the Center for Epidemiological Studies Depression (CES-D) Scale. This instrument has been widely used in dementia caregiving research, but the fit of this multifactor model and the explanatory contributions of multifactor models have not been sufficiently examined for caregiving samples. The authors subjected CES-D data ($N = 1,183$) from the initial Resources for Enhancing Alzheimer's Caregiver Health Study to confirmatory factor analysis methods and found that the 4-factor model provided excellent fit to the observed data. Invariance analyses suggested only minimal item-loading differences across race subgroups and supported the validity of race comparisons on the latent factors. Significant race differences were found on 3 of the 4 latent factors both before and after controlling for demographic covariates. African Americans reported less depressed affect and better well-being than White caregivers, who reported better well-being and fewer interpersonal problems than Hispanic caregivers. These findings clarify and extend previous studies of race differences in depression among diverse samples of dementia caregivers.

Keywords

caregiving; depression; race/ethnicity; confirmatory factor analysis; invariance

Providing care to a family member with Alzheimer's disease or another form of progressive dementia is often a chronically stressful experience that is associated with declines in physical health and psychological well-being (Aranda, 2001; Roth, Haley, Owen, Clay, & Goode, 2001; Schulz, O'Brien, Bookwala, & Fleissner, 1995). One of the most common problems found among dementia caregivers is emotional distress, and this is often measured by assessing the frequency or severity of symptoms of clinical depression. In addition to the large body of research showing elevated depression levels in numerous samples of caregivers compared with

noncaregivers (Pinquart & Sörensen, 2003), ethnic differences have also been reported in caregiver depression, stress reaction, and psychological well-being. However, different samples and measuring instruments have clouded the consistency and interpretability of many of these ethnic differences (Pinquart & Sörensen, 2005).

One of the most frequently used self-report instruments of depressive symptoms in caregiving research is the Center for Epidemiologic Studies Depression (CES-D) Scale. The CES-D consists of 20 items that inquire about the frequency of depressive symptoms during the week prior to the assessment (Radloff, 1977). Previous exploratory and confirmatory factor analysis studies have consistently found the 20-item CES-D to be a multidimensional instrument with as many as four correlated but distinct factors (Hertzog, Van Alstine, Usala, & Hultsch, 1990; Knight, Williams, McGee, & Olaman, 1997; Nguyen, Kitner-Triolo, Evans, & Zonderman, 2004; Radloff, 1977; Wong, 2000). The four underlying factors were labeled by Radloff (1977) as Depressed Affect, Somatic Symptoms, Positive Affect, and Interpersonal Problems. In several studies, confirmatory factor analysis methods have supported the validity of this four-factor model for the CES-D on data from diverse populations such as older adults (Hertzog et al., 1990; Knight et al., 1997), the homeless (Wong, 2000), low socioeconomic status African Americans (Nguyen et al., 2004), and Mexican Americans (Golding & Aneshensel, 1989). Other studies, however, have found that the depressed affect and somatic symptom items load onto a single factor, resulting in a three-factor model that provides sufficient fit to the data (Guarnaccia, Angel, & Worobey, 1989; Manson, Ackerson, Dick, Baron, & Fleming, 1990).

The majority of studies in dementia caregiving research that have used the CES-D have used only the CES-D total score (e.g., Beeson, Horton-Deutsch, Farran, & Neundorfer, 2000; Gallicchio, Siddiqi, Langenberg, & Baumgarten, 2002; Haley et al., 1996; Jensen, Ferrari & Cavanaugh, 2004; Roth et al., 2001; Schulz et al., 2004). To our knowledge, only one previous study has examined the factor structure of the CES-D in dementia caregivers, and that analysis found excellent fit for the four-factor model and minimal differences in factor structure when male and female caregivers were compared (O'Rourke, 2005). However, the relative advantages of the four-factor model over more parsimonious models with fewer factors and the fit of the four-factor model across multiple ethnic groups were not examined in the O'Rourke (2005) analysis.

Another limitation of the existing literature is the relative lack of psychometric studies on the equivalence of the CES-D items across different race or ethnic groups. Ethnic group comparisons on scale scores rely on an assumption that the measurement properties of the items on the scale are equivalent across those race subgroups (Bingenheimer, Raudenbush, Leventhal, & Brooks-Gunn, 2005; Stewart & Napoles-Springer, 2003). However, this assumption is usually not evaluated even though item-level data are typically available that would allow such an examination. Differential item functioning (DIF) analyses can be used to examine whether certain items are more or less sensitive indicators of an underlying latent construct for different subgroups. Methods exist within both confirmatory factor analysis (CFA) and item response theory (IRT) that allow researchers to distinguish DIF from true group differences on the underlying construct (Stark, Chernyshenko, & Drasgow, 2006). Although a few recent studies have examined DIF across race groups on CES-D items (e.g., Yang & Jones, 2007), to our knowledge no previous study has examined DIF for the CES-D items across multiple race or ethnic groups within the context of a multifactor model. Consequently, a systematic investigation of the psychometric properties of the CES-D items and the multifactor structure of this instrument across multiple race groups was conducted to support and inform ongoing research on racial differences in caregiver emotional distress.

In the present investigation, we sought (a) to compare the fit of competing factor models across African American, Hispanic, and White family caregivers, (b) to examine item and factor equivalence across these race groups, and (c) to compare caregivers from these race groups on the identified latent constructs underlying the CES-D items. We accomplished these objectives using baseline data from the initial Resources for Enhancing Alzheimer's Caregiver Health (REACH I) Study. The REACH I study was a multisite investigation of 1,229 caregiver-care recipient dyads from six locations across the United States (Gitlin et al., 2003; Wisniewski et al., 2003). The data are publically available through the Inter-University Consortium for Political and Social Research (ICPSR; Schulz, 2003). In the present analyses, CFA methods were used to compare the fit of one-, two-, three-, and four-factor models for the baseline CES-D data from all six sites. The one-factor model simply allowed all 20 items to load on a single latent construct. The two-factor model distinguished the 4 reverse-scored positive affect items into a separate Well-Being factor from the remaining 16 depressive symptom items. The design of the three-factor model elaborated on the two-factor model, with 2 items from the 16 depressive symptom items in the two-factor model designated as a third factor for interpersonal problems. The four-factor model was the same as that identified previously (e.g., Hertzog et al., 1990; Radloff, 1977) and included Depressed Affect (7 items), Well-Being (4 items), Interpersonal Problems (2 items), and Somatic Symptoms (7 items). Once the optimal measurement model was identified, we performed subsequent analyses to examine the equivalence of this factor structure across the three race groups and to determine whether the race groups differed on latent factor means after controlling for the effects of caregiver age, gender, education, and the caregiver-care recipient relationship.

Previous analyses of race differences in caregiver depression have been conducted as part of two articles in which baseline data from specific REACH I sites were used. Haley and colleagues (2004) found no differences between White and African American caregivers across four of the six sites on the total score of the CES-D but did find that African American caregivers reported better well-being than White caregivers on the four reverse-scored items that inquire about positive states. In addition, Coon and colleagues (2004) conducted comparisons between female Hispanic and female White caregivers at the Miami and Palo Alto sites and found no significant differences on the CES-D total score or the Well-Being subscale. Taken together, these two previous REACH I studies found minimal differences among race/ethnicity groups on caregiver depression. However, neither of the previous REACH I analyses incorporated a factor analysis of CES-D items, included data from all six sites, or examined differences across all three race groups simultaneously.

We conducted the present analyses to extend the previously reported REACH I findings by incorporating a latent variable measurement model and examining potential race differences in the context of a multidimensional measurement model. We hypothesized that the superior fit of the four-factor model over more parsimonious factor models would be replicated in this sample of dementia caregivers and that the four-factor model would provide adequate fit for all three race subgroups. We further hypothesized that race differences would emerge on confirmed latent factors that might otherwise be obscured by simple total score analyses. Such effects would be informative for isolating and refining the search for meaningful differences by race or ethnicity among family caregivers of persons with dementia.

Method

Participants

The participants in the REACH I Study were 1,229 primary family caregivers of persons with Alzheimer's disease or a related dementia. They were recruited for the purpose of developing interventions for caregivers of individuals with moderate cognitive impairment. More detailed information about the recruitment procedures and the interventions tested at each site have

been reported in previous articles (Coon, Schulz, & Ory, 1999; Gitlin et al., 2003; Schulz et al., 2004; Wisniewski et al., 2003). Only baseline data (before any interventions were initiated) were analyzed in the present article.

All data were sent to the coordinating center at the University of Pittsburgh, where they were verified for accuracy and organized into suitable files for analysis. The data were later made available for analysis by other researchers through the Inter-University Consortium for Political and Social Research (ICPSR) Web site (Schulz, 2003). All analyses reported in this article were based on item-level data downloaded from this Web site.

Of the 1,229 participants at baseline, 35 (2.8%) had missing values on some CES-D items, resulting in 1,194 cases with complete CES-D item data. An additional 11 caregivers (0.9%) could not be classified as African American, Hispanic, or White, resulting in 1,183 cases (96.3%) with complete CES-D item and race data for the measurement modeling analyses. The use of listwise deletion or complete cases only was considered acceptable in this case because only a small percentage of cases had partial or incomplete data (Tabachnick & Fidell, 2007).

The 1,183 participants included in the measurement models were recruited to participate across all six sites—Birmingham, Alabama: $n = 139$; Boston, Massachusetts: $n = 94$; Memphis, Tennessee: $n = 240$; Miami, Florida: $n = 214$; Palo Alto, California: $n = 247$; and Philadelphia, Pennsylvania: $n = 249$. The analysis sample consisted of 964 female (81.5%) and 219 male (18.5%) caregivers. Both genders were recruited at all sites with the exception of Palo Alto, where only female caregivers were recruited. The sample consisted of 681 White, 294 African American, and 208 Hispanic caregivers. Caregiver–care recipient relationships varied, with both 570 spouses (48.2%) and 613 (51.8%) nonspouses serving as caregivers. The nonspouses consisted mostly of adult children ($n = 494$), with the remainder ($n = 119$) being other family members.

The care recipients were community-dwelling older adults (528 men, 655 women) with either (a) a medical diagnosis of probable Alzheimer's disease or a related dementia or (b) a score of less than 24 on the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975).

Procedures and Measures

Research staff at each site screened potential participants by telephone using questions shared across sites. Caregivers who were eligible for the study and gave informed consent to participate were then administered a core battery of study measures during a follow-up visit, which typically took place in the caregiver's home. Each site obtained approval for all protocols and for the intervention that was specific to that site from its local institutional review board. The coordinating center at the University of Pittsburgh conducted periodic site visits to ensure adherence to study protocols and to verify the exclusive use of research interviewers who had been adequately trained and certified.

All measures were chosen on the basis of having established psychometric properties and being appropriate for an ethnically diverse sample (Switzer, Wisniewski, Belle, Dew, & Schulz, 1999). Study instruments were made available in both English and Spanish. The Spanish version was conceptually translated by a team that consisted of professional translators, bilingual/bicultural research staff, and an advisory board. The translations aimed at maximizing the meaning, intent, and understanding of the measures, and all materials were back-translated to English to confirm the consistency of the translations.

Demographic variables—Caregivers indicated their age, gender, primary racial or ethnic group (using census categories), and their relationship to the care recipient during the baseline

interview. Number of years of education was obtained, and yearly household income was recorded in one of 10 ordinal categories, ranging from less than \$5,000 per year to \$70,000 or more per year. Care recipients' general cognitive status was assessed using the MMSE (Folstein et al., 1975), a brief screening instrument that assesses orientation, memory, basic attention, language, and visuospatial/visuoconstructive abilities. Scores on the MMSE range from 0 to 30, with lower scores indicating greater cognitive impairment. In the present sample, care recipients represented a wide range of cognitive impairment as indicated by MMSE scores ranging from 0 to 29 ($M = 12.56$, $SD = 7.66$).

Depressive symptoms—Caregiver depressive symptoms were assessed using the 20 items from the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977). These items inquire about the frequency of depressive symptoms during the preceding week. Response options include 0 (*rarely or none of the time—less than 1 day per week*), 1 (*some or a little of the time—1–2 days per week*), 2 (*occasionally or a moderate amount of time—3–4 days per week*), or 3 (*most or almost all of the time—5–7 days per week*). The four Well-Being items are reverse scored when a total score is calculated. Higher total scores on the CES-D indicate greater depression. A score of 16 or more was originally considered suggestive of clinically significant depression (Radloff & Teri, 1986), although a higher cutpoint of 21 has been recommended for older individuals on the basis of an analysis of primary care patients (Lyneess, Noel, Cox, King, Conwell, & Caine, 1997).

Analysis Methods

The measurement models reported in this article were estimated and evaluated with the structural equation modeling procedures of the Mplus software package (Version 5.0; Muthén & Muthén, 2007). We conducted analyses of the different multifactor models across the entire sample of 1,183 participants and within each race group separately using weighted least squares estimation and mean- and variance-adjusted chi-square tests (WLSMV option in Mplus). This method attempts to recreate the observed polychoric correlation matrix for the CES-D items in the context of the model constraints. The degrees of freedom for the chi-square test of model fit are estimated empirically from the observed data and not directly determined from the model constraints when WLSMV estimation is used. All of the more parsimonious multifactor models were nested in the more complex models, and the statistical significance of the increments in fit of the one-, two-, three-, and four-factor models was tested using the DIFFTEST option as described by Muthén and Muthén (2007).

Once the optimal multifactor model was determined from these analyses, multiple group analyses were conducted to test the invariance of the factor loadings and the factor correlations across race groups. Testing factorial invariance for ordered-categorical items is more complicated and involves more options than when the observed item-level data are assumed to be continuous and normally distributed (Millsap & Tein, 2004). In the approach adopted here, the primary focus was on the invariance of the factor loadings. We constrained item thresholds to be equal across race groups, and we constrained the variances of the latent factors to be equal across groups by fixing all factor variances to 1. Factor means were fixed to 0 in the White group and were left free to vary in the African American and Hispanic groups.

Factor covariances, which were also factor correlations due to the fixing of the factor variances to 1 in each race group, were free to vary across the three race groups in the factor-loading invariance analyses. In the context of these specifications, the factor loadings were then either constrained to be equal across the race groups or unconstrained and allowed to take on separate estimates within each race group. Because we standardized the variance of the latent factors to 1 within each race group for both the constrained and unconstrained factor-loading models, these analyses tested whether within-group changes in standard deviation units on the latent

factors led to the same changes on the indicator variables across the three race groups. When examining constrained versus unconstrained factor correlations, we constrained the factor loadings to be equal across race groups. In both the factor-loading and factor-correlation models, the constrained models were nested within the respective unconstrained models, allowing a statistical test of invariance using the DIFFTEST procedure.

Measures of model fit included the chi-square goodness-of-fit test for each model, the comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA). Using both CFI and RMSEA, we assessed overall model fit while also taking into account the complexity or parsimony of the model. The CFI ranges from 0 to 1, with values greater than .95 considered indicative of excellent fit. The RMSEA assesses lack of fit, and values less than .05 or .06 are typically considered to indicate excellent fit to the observed data (Browne & Cudeck, 1993). In addition to chi-square tests of the statistical significance of improvements in fit for the less parsimonious models when conducting nested comparisons, we also used the CFI and RMSEA indices to assess any practical improvements in fit.

Differences in latent means across the three race groups were examined from both the constrained and unconstrained multiple group invariance analyses. However, the interpretation of such differences is limited by the fact that no covariates were included in the multiple group invariance analyses. If the results of the multiple group analyses indicate that the factor loadings are similar across the groups, then race group differences on the CES-D latent factors can also be examined with a multiple-indicator multiple-cause (MIMIC) modeling approach. The MIMIC modeling approach is more straightforward than a multiple group approach for testing the effects of grouping variables on latent score means after one has adjusted for the effects of other exogenous covariates (Thompson & Green, 2006). Consequently, a MIMIC model analysis was conducted that included race and four other exogenous predictors of the CES-D factor scores. In total, six exogenous predictor variables were examined including two coded contrasts for race (African American vs. White and Hispanic vs. White); the age of the caregiver in years; caregiver gender (female = 1, male = 0); whether the caregiver was a spouse (1) or some other family member (0); and caregiver education (number of years completed).

Results

Descriptive Information

Descriptive data for this sample of dementia caregivers are summarized in Table 1. One-way analyses of variance indicated significant group differences on caregiver age, education, and household income ($ps < .0001$). White caregivers were older, had more education, and higher incomes than African American and Hispanic caregivers, who tended to be more similar on these socio-demographic variables. Groups also differed significantly on care recipient MMSE scores, $F(2, 1149) = 14.50, p < .0001$, with African American care recipients having lower scores than White or Hispanic care recipients.

Chi-square tests indicated significant differences across race groups on care recipient gender and the caregiver-care recipient relationship ($ps < .0001$). African American caregivers had the highest proportion of female care recipients, and White caregivers were significantly more likely to be spouses than African American or Hispanic caregivers. In spite of the gender exclusion at the Palo Alto site, Hispanic caregivers only tended to be more likely to be female than the caregivers in the other two race group, $\chi^2(2, N = 1,183) = 5.65, p = .06$.

An analysis of variance revealed that the three race groups differed significantly on the CES-D total score, $F(2, 1180) = 16.42, p < .0001$. Pairwise comparisons using Scheffe's test revealed significant differences ($ps < .05$) between all three pairs of groups. Hispanic caregivers reported the highest levels of depressive symptoms, and African American caregivers reported the

lowest levels. Overall, 41% of the caregivers had total scores of 16 or greater on the CES-D, and 28% had scores of 21 or greater. The caregivers with total scores of 16 or greater included 31% of the African American, 55% of the Hispanic, and 41% of the White caregivers. The caregivers with total scores of 21 or greater included 20% of the African American, 40% of the Hispanic, and 27% of the White caregivers. We conducted further examinations of race differences in depressive symptoms after identifying the latent factors underlying the CES-D from our measurement models and after taking into account the effects of caregiver age, gender, education, and the caregiver–care recipient relationship as covariates in the MIMIC model.

Confirmatory Factor Analyses of the Multifactor Measurement Models

Table 2 summarizes the fit of the one-, two-, three-, and four-factor models using the WLSMV estimation method for the poly-choric correlation matrix of the 20 CES-D items across the entire sample of 1,183 participants. We found the simple one-factor model fit reasonably well, although nested comparisons that we performed using the DIFFTEST procedure indicated that all of the multifactor models fit significantly better than the one-factor model ($p < .0001$ for all comparisons). The four-factor model was found to fit the observed data significantly better than any of the nested and more simple multifactor models ($p < .0001$ for all comparisons), and the RMSEA and CFI statistics also indicated that the four-factor model provided the best fit among these alternative models. Overall, excellent fit was indicated for the four-factor model.

Similar results were obtained when the one-, two-, three-, and four-factor models were evaluated and compared within each race group separately. The nested comparisons again revealed that the four-factor model fit significantly better than any of the more parsimonious nested models ($p < .007$ for all comparisons). The RMSEA measure of fit for the four-factor model was strongest for the Hispanic group (.033), but was also generally in the acceptable range for the African American (.064) and White (.060) caregiving groups. Taken together, both total sample and individual group analyses supported the superior fit of the four-factor model over the more parsimonious alternative models, and the four-factor model was subsequently adopted as the optimal measurement model for the factor-loading invariance analyses.

Tests of Factor Invariance Across Race Groups for the Four-Factor Model

The factor loadings from the multiple group analyses for the four-factor model are displayed in Table 3. All factor loadings listed in Table 3 were significantly different from zero ($ps < .0001$), and none of the items were permitted to have nonzero factor loadings on any other factors besides the factor indicated in Table 3. The constrained factor loadings in Table 3 were obtained from the model that forced the factor loading for each item to be equivalent across the three race groups. This model was nested in the less parsimonious unconstrained model that allowed different loading estimates for each item in each race group. The comparison of these two nested models indicated that the unconstrained model fit significantly better than the constrained model, $\chi^2(29, N = 1183) = 56.37, p = .002$, but the fit indices were slightly better for the constrained model (CFI = .972, RMSEA = .056) than the unconstrained model (CFI = .958, RMSEA = .059). This suggests that the statistical improvements in fit were not of practical significance when weighed against the increases in model complexity for the completely unconstrained model.

Individual items were subsequently examined in a series of separate nested models in order to identify which specific items showed a significant lack of equivalence in factor loadings across the race groups. Because multiple comparisons were conducted, a more conservative Type I error rate of .01 was specified, consistent with the recommendations of Stark et al. (2006) for conducting multiple item-level analyses. As indicated in Table 3, only 3 of the 20 items showed

significant differences in loadings across the race groups. All 3 items were found to have slightly higher loadings on their respective factors in the White group compared with the other two race groups.

Table 4 presents the constrained and unconstrained factor correlations between the latent factors of the four-factor model across the three race groups. All factor correlations were significantly different from zero ($ps < .0001$). Moderate to high factor correlations were generally observed, and the depressed affect and somatic symptoms factors were particularly highly correlated in both the constrained and unconstrained models. This suggests that these two factors are only minimally distinct from each other. The nested comparison of the constrained and unconstrained factor correlation models indicated that the unconstrained model fit significantly better than the constrained model, $\chi^2(6, N = 1,183) = 18.61, p = .005$. The two fit indices, however, were quite similar in the constrained ($CFI = .979$, $RMSEA = .058$) and unconstrained ($CFI = .972$, $RMSEA = .056$) models. Individual factor correlations were sequentially tested in nested models to identify those correlations that showed statistically significant nonequivalence. The more conservative Type I error rate of .01 was again used for these multiple comparisons. Only the negative correlation between the Depressed Affect and Well-Being factors was found to vary significantly between race groups.

Group Differences on the CES-D Latent Factors

Because the results of the multiple group factor-loading invariance analyses indicated similar fit for the constrained and unconstrained models and only modest differences in factor loadings among the small number of items with significant group differences, we proceeded to examine race group differences on the CES-D latent factors using a MIMIC modeling approach. Exogenous variables were added to the single group four-factor measurement model and were examined as predictors of the CES-D latent factors. These exogenous predictors included caregiver age, gender, and education; caregiver relationship to the care recipient (i.e., spouse or nonspouse); and two coded vectors for caregiver race (African American vs. White and Hispanic vs. White contrasts).

Table 5 displays the effects of each predictor variable on the latent factor means in standard deviation units after the effects of all other predictors in the model have been statistically controlled. For example, an increase in age of 10 years from one caregiver to the next was associated with a decrease of -0.14 standard deviation units on the depressed affect latent factor ($10 * -0.014$), and female caregivers were found to be 0.34 standard deviation units higher on this latent factor than male caregivers.

The effects of the two race group contrasts are of primary interest from these analyses. After controlling for the effects of the other predictors in the model, we found that African American caregivers reported significantly less depressed affect and greater well-being than White caregivers, whereas Hispanic caregivers reported less well-being and more interpersonal problems than White caregivers. There were no statistically significant racial differences on the Somatic Symptoms factor.

Other statistically significant covariate-adjusted effects included findings that female caregivers reported more depressed affect and more somatic symptoms than male caregivers, that spouse caregivers reported more depressed affect and less well-being than nonspouse caregivers, and that both caregiver age and education were associated with fewer problems on three of the four CES-D latent factor domains.

Discussion

The results of this study indicate that the four-factor model of CES-D item responses provided excellent fit to the data obtained from a large multisite sample of family caregivers of individuals with dementia. The four-factor model was originally reported by Radloff (1977) using exploratory factor analysis procedures and was later supported by Hertzog and colleagues (1990) using CFA techniques on data from a sample of community-dwelling older adults. The excellent fit of the four-factor model to the data from the REACH I sample of dementia caregivers supports the measurement validity of this instrument in caregiving research and suggests that examinations of caregiver depression could be supplemented with more specific analyses on distinct clusters of depressive symptoms. It is interesting that the positively worded well-being items consistently loaded on a separate factor, a finding that has also been observed in factor analyses of the 15-item Geriatric Depression Scale (Brown, Woods, & Storandt, 2007). It would be helpful to know for both clinical and research purposes, for example, whether an older adult's overall depression score is largely due to a tendency to endorse affective or somatic symptoms or to deny positively worded well-being items.

Our findings are consistent with the results reported by O'Rourke (2005), who also found excellent fit for the four-factor model of CES-D item responses for a large sample of dementia caregivers. The present analyses extended beyond O'Rourke's findings in that (a) the four-factor model was shown to fit significantly better than more parsimonious models with fewer factors and (b) the differences in factor structure and latent factor means by race subgroups were examined. Confirmation of the validity of these subscales in commonly used instruments for assessing depression in older adults allows for more fine-grained analyses of the underlying components of depression.

The multiple group analyses indicated that the four-factor model fits quite well across the all three race groups in the REACH I database. The fit indices revealed that models with invariant or constrained factor loadings and factor correlations across the race groups fit the data as well as comparable unconstrained models when model parsimony was also taken into account. Although 3 of the 20 factor loadings and one of the six factor correlations showed statistically significant differences across the three race groups, the general patterns of the fit statistics, item loadings, and factor correlations were remarkably similar across the groups. Nested comparisons suggested only small and largely trivial improvements in fit when unconstrained measurement models were estimated. As described by Millsap (2005), *complete factorial invariance* is easy to define in theory but rarely encountered in practice, especially with large samples, and the pattern of the present results is consistent with Millsap's interpretive label of *approximate invariance*. In these instances, the degree of nonequivalence in group-specific estimates is deemed to be minimal according to overall model fit standards, and this minimal nonequivalence is not likely to lead to notable biases in the standard use of the instrument in practice.

The factor-loading invariance analyses that were conducted in the context of our multiple group CFA models represent one standard approach for examining DIF, that is, whether individual items are more or less sensitive indicators for detecting differences in underlying latent constructs for some groups of respondents over others. These questions are typically examined using either IRT or multiple group CFA approaches. These methods address similar questions, although they traditionally proceed via different procedures and assumptions including unidimensional versus multidimensional latent construct models, assumptions of linear or curvilinear relationships between latent constructs and observed item responses, and constrained versus unconstrained models as the baseline for individual item invariance comparisons (Bingenheimer et al., 2005; Stark et al., 2006). Recent progress has been made in reconciling the two approaches, especially for unidimensional models in which nonlinear

estimation procedures are used in CFA approaches that assume ordinal item-response data (e.g., Yang & Jones, 2007). We extended these innovations by analyzing nonlinear ordinal item-response data in the context of a multifactor, multiple group CFA model. Additional methodological work is needed to further advance these CFA-based methods and to compare them with IRT approaches. Millsap and Tein (2004) have described the complexities of extending multiple group invariance analyses to CFA approaches with ordinal data, and Stark and colleagues (2006) have pointed out that complex multidimensional models are only in the early stages of development from the IRT tradition. Continued progress in these areas is likely to be forthcoming in the next few years and might benefit from systematic Monte Carlo simulation studies.

At present, these results do suggest that ethnic group comparisons on subscales of depressive symptoms would be valid for dementia caregivers. Such comparisons could facilitate better understanding of the nature of ethnic group differences in caregiver emotional distress. The results of our MIMIC model that examined differences on the latent factors as a function of caregiver race and other demographic variables confirmed that these dimensions were differentially sensitive to meaningful group differences. After adjusting for age, gender, education, and relationship differences, we found significant differences by race on three of the four latent factors.

It is noteworthy that no race group differences were found on the Somatic Symptoms factor after adjustments were made for model covariates. A large number of items load on the Somatic Symptoms factor, and the content of these items appears to be more heterogeneous than the content of items on the other three factors. In addition to somatic complaints about appetite and sleep problems, this factor also includes items addressing concentration problems and poor energy levels. Although this factor was highly correlated with the Depressed Affect factor, a critical distinction between these two latent constructs was that the Depressed Affect factor did detect significant differences between African American caregivers and the other two race groups.

The greater well-being of the African American caregivers is consistent with the earlier findings of Haley and colleagues (2004), who analyzed raw subscale scores from the Birmingham, Boston, Memphis, and Philadelphia sites of REACH I and found that African American caregivers reported greater well-being than White caregivers. Other baseline analyses of REACH I data have shown that, compared with White caregivers, African American caregivers reported lower appraisals of stress in reaction to care recipient memory and behavior problems (Roth et al., 2003), lower upset ratings in response to problems with both basic and instrumental activities of daily living (Gitlin et al., 2005), and higher positive aspects of caregiving (Roff et al., 2004). These results are generally consistent with other studies of family caregivers that have found African Americans to report less depression and greater resilience in the caregiving role (Connell & Gibson, 1997). Meta-analysis methods, however, have shown that these effects can vary considerably depending on whether convenience sampling methods are used or whether spouse or nonspouse caregivers are compared across race groups (Pinquart & Sörensen, 2005). The absence of a noncaregiving comparison group in the REACH I study also makes it difficult to determine whether the race differences found in the present analyses are unique to family caregivers of persons with dementia or if similar race differences might be detected in more general populations of middle-aged and older adults. However, comparisons of African American and White family caregivers and noncaregivers from previous studies suggest that racial differences in depression are stronger in caregiving groups than in noncaregiving comparison groups (Haley et al., 1996; Roth et al., 2001). The present analyses contribute information that further clarify these previous effects by suggesting that ethnically diverse groups do not differ on the more somatic

or energy-related symptoms of depression and that African American and White caregivers do not differ on the interpersonal symptoms dimension.

Previous comparisons of raw CES-D scores for the Hispanic and White participants at the Miami and Palo Alto sites revealed no racial differences (Coon et al., 2004), but those analyses did not take into account data from White caregivers at other REACH I sites, did not take caregiver relationship into account, and did not examine factor analytically defined CES-D subscales. The present analyses that were based on the entire REACH I sample and used CES-D latent variables as outcomes appeared to be more sensitive and informative for identifying areas in which greater problems were reported by the Hispanic caregivers in comparison with the other two racial groups. The symptom clusters identified by the latent variables, therefore, appear to provide greater clarity on the nature of depressive symptom differences among these three race groups.

Additional findings from the MIMIC model revealed that after adjustments were made for other predictors in the model, older caregivers reported less depressed affect and fewer interpersonal problems and somatic symptoms than did younger caregivers; women reported more depressed affect and more somatic symptoms than men; and education was associated with fewer symptoms of depression on three of the four latent factors. Furthermore, spouse-caregivers reported more depressed affect and less well-being than other family members serving in the caregiving role. Dementia caregiving can be particularly stressful for spouses because the caregivers are not only managing the considerable burdens of providing care but are also experiencing significant losses in social support and other important aspects of the marital relationship.

Future research should examine whether intervention-induced reductions in depression and general improvements in emotional well-being are detected across all CES-D subscales or are limited to one or more of these dimensions of affective functioning. An intervention designed to strengthen the social support resources of spouses of persons with dementia has been shown to reduce depressive symptoms as measured by the Geriatric Depression Scale (Mittelman, Roth, Coon, & Haley, 2004), but additional research is needed to further examine the nature of these improvements. By examining differences on validated subscales, researchers should be able to test whether spouse and nonspouse caregivers show different trajectories of improvement on different dimensions of depressive symptoms. The differences between spouse and nonspouse caregivers in the present analyses suggest that some latent factors of depressive symptoms might be more sensitive than other factors to intervention-induced improvements for certain subgroups of caregivers. Spouse caregivers, for example, might benefit more than nonspouses from interventions that are targeted at improving feelings of well-being and reducing affective symptoms. In general, it is likely that the pattern of intervention effects will depend on the sample being studied and the specific targets of the intervention program.

There are a number of limitations of the present analyses that should be acknowledged and addressed in future research. Each site used somewhat different sampling and recruitment procedures, and there is substantial overlap between ethnic group membership and geographical site in the REACH I sample. These considerations warrant caution in generalizing the present findings to the larger national population of ethnically diverse caregivers, and firm generalizations must await replications with data from more representative national samples. We considered conducting separate factor analysis models for each site, but sample size considerations rendered such an approach impractical, especially for examining possible racial differences within each site. Our approach of first confirming the superior fit of the four-factor model compared with simpler competing models, then evaluating the invariance of the factor structure across race groups, and then examining race differences on confirmed latent factors

using a MIMIC modeling approach was successful for demonstrating the appropriateness of the four-factor model and for identifying covariate-adjusted race differences on the latent factors of depressive symptoms. However, it is possible that differences in recruitment strategies across sites and other potential geographic differences could have contributed to the observed race differences in these analyses. Many previous studies of racial differences among family caregivers have been limited to single geographic locations, and the REACH studies represent a step forward in that data are collected from caregivers at multiple sites across the country. Nonetheless, additional work is necessary to further examine the geographic and cultural extent of these race differences across multiple locations and socioeconomic classifications.

At present, it is well established that depression is a common problem among family caregivers of persons with dementia. The CES-D is widely used to assess the presence of depressive symptoms in this population, and the original four-factor model of item responses is informative for identifying meaningful clusters of depressive symptoms in dementia caregivers. The four-factor model was effective for identifying significant differences among African American, Hispanic, and White caregivers, and more focused analyses of the multiple dimensions of depressive symptoms should be conducted in future research. These analyses could be useful for better characterizing the nature of problems experienced by particular subgroups of caregivers and for tracking the precise improvements that are achieved with targeted intervention programs.

References

- Aranda MP. Racial and ethnic factors in dementia caregiving research in the U.S. *Aging and Mental Health* 2001;5:S116–S123. [PubMed: 11513487]
- Beeson R, Horton-Deutsch S, Farran C, Neundorfer M. Loneliness and depression in caregivers of persons with Alzheimer's disease or related disorders. *Issues in Mental Health Nursing* 2000;21:779–806. [PubMed: 11854982]
- Bingenheimer JB, Raudenbush SW, Leventhal T, Brooks-Gunn J. Measurement equivalence and differential item functioning in family psychology. *Journal of Family Psychology* 2005;19:441–455. [PubMed: 16221024]
- Brown PJ, Woods CM, Storandt M. Model stability of the 15-item Geriatric Depression Scale across cognitive impairment and severe depression. *Psychology and Aging* 2007;22:372–379. [PubMed: 17563192]
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Newbury Park, CA: Sage; 1993. p. 136-162.
- Connell CM, Gibson GD. Racial, ethnic, and cultural differences in dementia caregiving: Review and analysis. *The Gerontologist* 1997;37:355–364. [PubMed: 9203759]
- Coon DW, Rubert M, Solano N, Mausbach B, Kraemer H, Arguelles T, et al. Well-being, appraisal, and coping in Latina and Caucasian female dementia caregivers: Findings from the REACH study. *Aging and Mental Health* 2004;8:330–345. [PubMed: 15370049]
- Coon, DW.; Schulz, R.; Ory, MG. Innovative intervention approaches with Alzheimer's disease caregivers. In: Biegel, D.; Blum, A., editors. *Innovations in practice and service delivery across the lifespan*. New York: Oxford University Press; 1999. p. 295-325.
- Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;12:189–198. [PubMed: 1202204]
- Gallicchio L, Siddiqi N, Langenberg P, Baumgarten M. Gender differences in burden and depression among informal caregivers of demented elders in the community. *International Journal of Geriatric Psychiatry* 2002;17:154–163. [PubMed: 11813279]
- Gitlin LN, Belle SH, Burgio L, Czaja SJ, Mahoney D, Gallagher-Thompson D, et al. Effect of multicomponent interventions on caregiver burden and depression: The REACH multisite initiative at 6-month follow-up. *Psychology and Aging* 2003;18:361–374. [PubMed: 14518800]

- Gitlin LN, Roth DL, Burgio LD, Loewenstein DA, Winter L, Nichols L, et al. Caregiver appraisals of functional dependence in individuals with dementia and associated caregiver upset. *Journal of Aging and Health* 2005;17:148–171. [PubMed: 15750049]
- Golding JM, Aneshensel CS. Factor structure of the Center for Epidemiologic Studies Depression Scale (CES-D) among Mexican Americans and non-Hispanic whites. *Psychological Assessment* 1989;1:163–168.
- Guarnaccia PJ, Angel R, Worobey IL. The factor structure of the CES-D in the Hispanic health and nutrition examination survey: The influences of ethnicity, gender, and language. *Social Science and Medicine* 1989;29:85–94. [PubMed: 2740931]
- Haley WE, Gitlin LN, Wisniewski SR, Mohoney DF, Coon DW, Winter L, et al. Well-being, appraisal, and coping in African-American and Caucasian female dementia caregivers: Findings from the REACH study. *Aging and Mental Health* 2004;8:316–329. [PubMed: 15370048]
- Haley WE, Roth DL, Coleton MI, Ford GR, West CA, Collins RP, et al. Appraisal, coping, and social support as mediators of well-being in Black and White family caregivers of patients with Alzheimer's disease. *Journal of Consulting and Clinical Psychology* 1996;64:121–129. [PubMed: 8907091]
- Hertzog C, Van Alstine J, Usala PD, Hultsch DF. Measurement properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations. *Psychological Assessment* 1990;2:64–72.
- Jensen CJ, Ferrari M, Cavanaugh JC. Building on the benefits: Assessing satisfaction and well-being in elder care. *Ageing International* 2004;29:88–110.
- Knight RG, Williams S, McGee R, Olanoff S. Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. *Behavioral Research Therapy* 1997;35:373–380.
- Lyness JM, Noel TK, Cox C, King DA, Conwell Y, Caine ED. Screening for depression in elderly primary care patients. *Archives of Internal Medicine* 1997;157:449–454. [PubMed: 9046897]
- Manson SM, Ackerson LM, Dick RW, Baron AE, Fleming CM. Depressive symptoms among American Indian adolescents: Psychometric characteristics of the Center for Epidemiologic Studies Depression Scale (CES-D). *Psychological Assessment* 1990;2:231–237.
- Millsap, RE. Four unresolved problems in studies of factorial invariance. In: Maydeu-Olivares, A.; McArdle, JJ., editors. *Psychometrics: A festschrift to Roderick P. McDonald*. Mahwah, NJ: Erlbaum; 2005. p. 153–171.
- Millsap RE, Tein JY. Assessing factorial invariance in ordered categorical measures. *Multivariate Behavioral Research* 2004;39:479–515.
- Mittelman MS, Roth DL, Coon DW, Haley WE. Sustained benefit of supportive intervention for depressive symptoms in Alzheimer's caregivers. *American Journal of Psychiatry* 2004;161:850–856. [PubMed: 15121650]
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. 4. Los Angeles, CA: Authors; 2007.
- 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. 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.
- Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: A meta-analysis. *Psychology and Aging* 2003;18:250–267. [PubMed: 12825775]
- Pinquart M, Sörensen S. Ethnic differences in stressors, resources, and psychological outcomes of family caregiving: A meta-analysis. *The Gerontologist* 2005;45:90–106. [PubMed: 15695420]
- 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;5:119–137.
- Roff LL, Burgio LD, Gitlin L, Nichols L, Chaplin W, Hardin JM. Positive aspects of Alzheimer's caregiving: The role of race. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 2004;59B:P185–P190.

- Roth DL, Burgio LD, Gitlin LN, Gallagher-Thompson D, Coon DW, Belle SH, et al. Psychometric analysis of the Revised Memory and Behavior Problems Checklist: Factor structure of occurrence and reaction ratings. *Psychology and Aging* 2003;18:906–915. [PubMed: 14692875]
- Roth DL, Haley WE, Owen JE, Clay O, Goode KT. Latent growth models of the longitudinal effects of dementia caregiving: A comparison of African American and White family caregivers. *Psychology and Aging* 2001;16:427–436. [PubMed: 11554521]
- Schulz R. Resources for Enhancing Alzheimer's Caregiver Health (REACH), [Computer file, ICPSR version]. Ann Arbor, MI: Inter-University Consortium for Political and Social Research (ICPSR); 2003. Available from <http://www.icpsr.umich.edu>
- Schulz R, Belle SH, Czaja SJ, McGinnis KA, Stevens AB, Zhang S. Long-term care placement of dementia patients and caregiver health and well-being. *Journal of the American Medical Association* 2004;292:961–967. [PubMed: 15328328]
- 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]
- Stark S, Chernyshenko OS, Drasgow F. Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology* 2006;91:1292–1306. [PubMed: 17100485]
- Stewart AL, Napoles-Springer AM. Advancing health disparities research. Can we afford to ignore measurement issues? *Medical Care* 2003;41:1207–1220. [PubMed: 14583684]
- Switzer GE, Wisniewski SR, Belle SH, Dew MA, Schulz R. Selecting, developing, and evaluating research instruments. *Social Psychiatry and Psychiatric Epidemiology* 1999;34:399–409. [PubMed: 10501709]
- Tabachnick, BG.; Fidell, LS. Using multivariate statistics. 5. New York: Harper Collins; 2007.
- Thompson, MS.; Green, SB. Evaluating between-group differences in latent variable means. In: Hancock, GR.; Mueller, RD., editors. *Structural equation modeling: A second course*. Greenwich, CT: Information Age; 2006. p. 119-169.
- Wisniewski SR, Belle SH, Coon DW, Marcus SM, Ory MG, Burgio L, et al. Resources for Enhancing Alzheimer's Caregiver Health (REACH): Project design and baseline characteristics. *Psychology and Aging* 2003;18:375–384. [PubMed: 14518801]
- Wong YI. Measurement properties of the Center for Epidemiologic Studies–Depression Scale in a homeless population. *Psychological Assessment* 2000;12:69–76. [PubMed: 10752365]
- Yang FM, Jones RN. Center for Epidemiologic Studies–Depression Scale (CES-D) item response bias found with Mantel–Haenszel method was successfully replicated using latent variable modeling. *Journal of Clinical Epidemiology* 2007;60:1195–1200. [PubMed: 17938063]

Table 1
Descriptive Information for the Three Race Groups and the Total Sample in the
Initial Resources for Enhancing Alzheimer's Caregiver Health (REACH I) Study

Variable	African American (N = 294)	Hispanic (N = 208)	White (N = 681)	Total (N = 1,183)
Caregiver age (years)				
M	58.2	58.9	65.0	62.2
SD	12.9	14.3	13.1	13.6
Caregiver gender (% female)	81.6	87.1	79.7	81.5
Caregiver education (years)				
M	12.78	11.61	13.66	13.08
SD	2.72	4.34	2.75	3.17
Caregiver household income (median)	\$18,718	\$19,545	\$30,370	\$26,349
Care recipient age (years)				
M	80.4	78.6	79.0	79.3
SD	10.3	12.0	8.2	9.5
Care recipient gender (% female)	69.7	58.7	48.2	55.4
Care recipient MMSE score				
M	10.49	12.93	13.35	12.57
SD	6.24	7.58	8.08	7.66
Caregiver relationship to care recipient				
Spouse (%)	25.9	40.4	60.2	48.2
Child (%)	55.4	49.5	33.5	41.8
Other (%)	18.7	10.1	6.3	10.1
Caregiver CES-D total score				
M	12.87	18.62	15.24	15.24
SD	10.27	11.55	11.24	11.21

Note. MMSE = Mini-Mental State Examination; CES-D = Center for Epidemiological Studies Depression Scale.

Table 2
Fit Statistics From the Weighted Least Squares Mean- and Variance-Adjusted Analyses of the CES-D Factor Models for the Entire Sample

Model	χ^2	df	p	CFI	RMSEA
One-factor	1281.07	101	<.0001	.848	.099
Two-factor	763.21	103	<.0001	.915	.074
Three-factor	591.87	104	<.0001	.937	.063
Four-factor	414.34	102	<.0001	.960	.051

Note. $N = 1,183$. CES-D = Center for Epidemiological Studies Depression Scale; CFI = comparative fit index; RMSEA = root-mean-square error of approximation.

Table 3
Factor Loadings for the Four-Factor Model From Constrained and Unconstrained Multiple Group Analyses

CES-D item	Factor	Constrained	Unconstrained		
			African American	Hispanic	White
Bothered by things	Somatic Symptoms	.61	.65	.61	.59
Appetite was poor	Somatic Symptoms	.57	.55	.48	.61
Can't shake the blues	Depressed Affect	.86	.81	.82	.88
Just as good as others	Well-Being	.49	.48	.37	.52
Trouble concentrating	Somatic Symptoms	.58	.62	.56	.57
Felt depressed	Depressed Affect	.88	.86	.87	.89
Everything was an effort ^a	Somatic Symptoms	.78	.66	.71	.84
Hopeful about the future ^a	Well-Being	.61	.44	.52	.67
Life has been a failure	Depressed Affect	.73	.79	.62	.73
Fearful ^a	Depressed Affect	.70	.65	.52	.76
Sleep was restless	Somatic Symptoms	.60	.63	.59	.59
Happy	Well-Being	.89	.89	.93	.88
Talked less than usual	Somatic Symptoms	.51	.57	.41	.50
Lonely	Depressed Affect	.79	.80	.78	.78
People were unfriendly	Interpersonal Problems	.77	.83	.73	.74
Enjoyed life	Well-Being	.86	.83	.82	.86
Crying spells	Depressed Affect	.74	.75	.75	.73
Sad	Depressed Affect	.86	.86	.87	.85
People disliked me	Interpersonal Problems	.89	.92	.94	.80
Could not get going	Somatic Symptoms	.67	.65	.71	.66

Note. CES-D = Center for Epidemiological Studies Depression Scale.

^a Item with loadings that are significantly different across the race groups in the unconstrained model, $p < .01$.

Table 4
Factor Correlations for the Four-Factor Model From Constrained and Unconstrained Multiple Group Analyses

Factor A	Factor B	Constrained	Unconstrained		
			African American	Hispanic	White
Depressed Affect	Well-Being ^a	−0.75	−0.71	−0.53	−0.82
	Interpersonal Problems	0.61	0.59	0.70	0.57
Well-Being	Somatic Symptoms	0.86	0.89	0.81	0.85
	Interpersonal Problems	−0.41	−0.37	−0.38	−0.44
Interpersonal Problems	Somatic Symptoms	−0.62	−0.62	−0.42	−0.68
		0.53	0.38	0.68	0.52

^aFactor combination with correlations that differ significantly across race groups in the unconstrained model, $p < .01$.

Table 5

Effects (in Standard Deviation Units) of Demographic Characteristics on the CES-D Latent Factors

Characteristic	Depressed affect	Well-being	Interpersonal problems	Somatic symptoms
Caregiver age (years)	-0.014***	0.002	-0.017***	-0.013***
Caregiver gender: female vs. male	0.341***	-0.136	-0.132	0.365***
Caregiver education (years)	-0.033***	0.011	-0.029*	-0.037**
Relationship: spouse vs. other	0.289***	-0.325***	-0.063	0.028
Race/ethnicity:				
African American. vs. White	-0.299***	0.427***	0.182	-0.118
Hispanic vs. White	0.147	-0.431***	0.407***	-0.080

Note. Significance tests were conducted by calculation of a *z* statistic from the estimate divided by its standard error. CES-D = Center for Epidemiological Studies Depression Scale.

*
 $p < .05$.

**
 $p < .01$.

 $p < .001$