

Confirmatory Factor Analysis of a Brief Version of the Zarit Burden Interview in Black and White Dementia Caregivers

Crystal V. Flynn Longmire, PhD^{*1} and Bob G. Knight, PhD²

¹Department of Neurosciences, Medical University of South Carolina, Charleston.

²Davis School of Gerontology and Department of Psychology, University of Southern California, Los Angeles.

*Address correspondence to Crystal V. Flynn Longmire, PhD, Medical University of South Carolina, 5900 Core Road, Suite 203, North Charleston, SC 29406. E-mail: flynnlo@musc.edu

Received October 7, 2010; Accepted January 26, 2011
Decision Editor: William J. McAuley, PhD

Purpose of the study: Although the Zarit Burden Interview (ZBI) is one of the most extensively used measures in research for caregiver burden, few researchers have examined its factor structure. Furthermore, though the ZBI has also been used in cross-group comparisons of burden, there have not been studies of whether or not it measures burden equally across various groups. Therefore, this study considers the psychometric properties of a brief version of the ZBI with particular attention to its factor structure and metric equivalence across two racial groups. **Design and Methods:** Distribution, reliability, and confirmatory factor analyses were performed with a 14-item three-factor ZBI factor model in a sample of 175 Black and 225 White caregivers of family members with dementia. **Results:** The 14 ZBI items were reliable and fairly normally distributed for both groups. The three-factor model fits the data and was invariant across the Black and White caregivers for number of factors, factor loadings, and factor covariances. **Implications:** These findings contribute to the literature on the factor structure of the ZBI and provide new data on the invariance of the ZBI across two racial/ethnic groups of caregivers. This study provides support for the validity of findings that compare the burden scores of Black and White caregivers in studies utilizing the ZBI. The 14-item version also offers a more parsimonious way to measure burden in clinical settings, potentially increasing screening opportunities when caregiver contact time is limited.

Key Words: Factor structure, Metric equivalence, Invariance, Racial/ethnic comparisons

As older adults live longer, they often have chronic conditions and a need for assistance (Institute of Medicine, 2007). Thus, as the number of older adults increase, the number of caregivers is also increasing. The idea that informal care of older adults with chronic illnesses, for example, dementia, often places a major burden on family members and other caregivers has been recognized for at least thirty years (Zarit, Reever, & Bach-Peterson, 1980). As such, there has been an emphasis in caregiver literature to understand the psychological and physical effects of caregiving and to assess the needs of caregivers in order to help them shoulder the burden of caregiving (Mohamed, Rosenheck, Lyketsos, & Schneider, 2010; Schulz, O'Brien, Bookwala, & Fleissner, 1995). It benefits everyone for caregivers to experience less negative effects such as burden, depression, or physical disability due to caring for relatives and friends. Benefits of less caregiver burden include better caregiver quality of life and care recipient health status and fewer institutional costs (Bell, Araki, & Neumann, 2001; Miller, Rosenheck, & Schneider, 2010). Given the increasing population of older adults, caregiver burden affects not only families, but society as a whole.

While there is an interest in the well-being of caregivers in general, there is also an understanding that groups of caregivers differ in background, resources, and care responsibilities among other things. As a result, their experiences and needs also may vary (Pinquart & Sörensen, 2005). Differences between Black and White caregivers have been one of the more common topics in ethnic caregiver research (Janevic & Connell, 2001). For instance, researchers have examined whether the effects of the stress of caregiving vary between Black caregivers and White caregivers. Specifically, when burden has been examined for differences by race group, Black caregivers were found to report less burden than White caregivers (Kosberg, Kaufman, Burgio, Leeper, & Sun, 2007; Lawton, Rajagopal, Brody, & Kleban, 1992; Pinquart & Sörensen, 2005), even when providing more care to more impaired older adults (Allen-Kelsey, 1998; Fredman, Daly, & Lazur, 1995).

Measurement Equivalence in Comparative Research

When comparing groups, it is fundamental to the research that the measures employed assess the same construct between comparison groups. It is also essential in applied settings that client evaluations operate as expected and in a similar manner for everyone. Meaningful, valid results can be obtained only when equivalent measures are used. If not, findings of group differences may actually be artifacts of differences in measurement (Tran, 1997). In regards to caregivers, scales that are not valid across groups can lead to erroneous study results and client assessments as well as ineffective policy and interventions.

Measurement equivalence (invariance) in comparative research has been discussed as having three forms: conceptual equivalence, metric equivalence, and structural equivalence (Tran, 1997). Conceptual equivalence refers to the sameness of meaning of research concepts across comparison groups. Metric equivalence assumes conceptual equivalence and means that observed items have comparable relationships with their respective latent concepts or factors across groups. Structural equivalence, assessed after conceptual and metric equivalence, relates to the sameness of causal relationships of a research concept for comparative groups. Metric equivalence was the objective of this analysis, using confirmatory factor analysis (Confirmatory factor analysis [CFA]) to test the

hypothesized factor structure of a brief version of a burden scale in Black and White caregivers.

Zarit Burden Interview

The present study focuses on the psychometric properties of the Zarit Burden Interview (ZBI; Zarit et al., 1980). The version of ZBI seen most often in the literature has 22 items. Twenty-one of the items are supposed to measure several aspects of burden, whereas Item 22 is a global measure of burden and is not usually included in factor analyses. The ZBI was developed to measure burden among caregivers of community-dwelling persons with dementia (Bachner & O'Rourke, 2007).

Despite the fact that the ZBI is one of the most extensively used measures in research for caregiver burden, few researchers have examined its factor structure. The originators of the ZBI did not provide a factor structure. Though the ZBI factor analysis literature mostly contains studies of brief versions of the ZBI, which commonly have two or three factors (O'Rourke & Tuokko, 2003b), the number of factors proposed in the literature for the ZBI has ranged from one to five (Hébert, Bravo, & Prévile, 2000). The ZBI has also been translated into other languages and tested in other cultures (Bachner & Ayalon, 2010; Kumamoto & Arai, 2004), and its reliability across populations of caregivers and patients has been examined (Bachner & O'Rourke, 2007). For the sake of simplicity and to keep the concepts consistent, our literature review focused on the two- and three-factor models in the North American/Western culture (U.S. and Canadian) studies of the ZBI as they were most applicable to our analyses of the English version. As with the vast majority of studies including the ZBI, and even more so those studying factor structure, the caregiver group of focus is dementia caregivers.

Hébert and colleagues (2000) developed a two-factor 12-item version of the ZBI. Using CFA, they tested several alternative models for fit before arriving at their final model. Their final result was that a version of a model originally proposed by Whitlatch, Zarit, and von Eye (1991), respecified via exploratory factor analysis (EFA) and modification index information with two factors and 12 items, was the best at depicting the burden reported by the caregivers in their sample ($\chi^2 = 64.75$, $df = 49$, $p = .07$). Other fit indices supported this conclusion: adjusted goodness-of fit index (AGFI) = .98; $\chi^2/df = 1.32$; root mean square residual (RMR) = .10. Their two

factors, like those of Whitlatch and colleagues, were called Personal Strain (three items) and Role Strain (nine items); however, the number of ZBI items included and the composition of the factors were not the same between the two studies. Hébert and colleagues' personal strain factor focused more on the negative feeling the caregiver may have regarding her caregiving responsibilities (e.g., Do you feel strained when you are around your relative?). Their role strain factor had items that ask about the effect caring has had on the caregiver's health and other relationships (e.g., Do you feel that your social life (health) has suffered because you are caring for (because of your involvement with) your relative?) Additionally, the Hébert model includes Item 22, the global item, which is not usually included in factor analysis.

Knight, Fox, and Chou (2000) confirmed a three-factor model twice in two different caregiver samples. They used CFA to test both 18-item and 21-item versions of the scale (Item 22—the global item: “Overall, how burdened do you feel in caring for your relative?”—was not included). Given that one- and two-factor models did not fit the data, they did an EFA using oblique rotation and found a three-factor model. Their three factors were Embarrassment/Anger (eight items; e.g., feel strained, have less privacy, feel social life has suffered, feel embarrassed), Patient's Dependency (four items; e.g., relative is dependent, not enough time for self), and Self-criticism (two items; should be doing more, could be doing a better job). This three-factor model, like all the other models in the literature, does not include all the 21 nonglobal items but 14 of them. The $\chi^2/df = 2.76$ and the goodness-of-fit index showed acceptable fit with a comparative fit index (CFI) = .91. One item “social life has suffered due to caring” loaded highly on both Embarrassment/Anger and Patient's Dependency. This item was placed on the Patient's Dependency factor, and the model was confirmed in a second sample of caregivers with a chi-square of 102.71, $df = 75$, and $p = .01$ and CFI = .94. Though not specifically mentioned, their χ^2/df showed good fit too because the ratio was less than two (1.37). Knight and colleagues concluded that the finding of a three-factor model having the best fit of the models tested is evidence that the ZBI is not unidimensional but does indeed have several dimensions just as other burden scales are multidimensional (Lawton, Kleban, Moss, Rovine, & Glicksman, 1989; Novak & Guest, 1989).

As mentioned earlier, Hébert and colleagues (2000) tested several models when developing theirs and one of them was the Knight and colleagues model. For their Canadian sample, the Knight model had a chi-square of 230.80, $df = 75$, $p < .01$, and the fit statistics were as follows: AGFI = .94, $\chi^2/df = 3.08$, and RMR = .26. The AGFI supported model fit, but the χ^2/df was borderline and the RMR should be approximately .05 or less in a well-fitting model (Byrne, 2001). The RMR here can be interpreted as meaning the Knight model explained the correlations to within the average error of .26. This is outside the desired range but so was the final Hébert model with an RMR of .10.

The Hébert and Knight models were later analyzed in another study. Using a nationally representative sample in Canada, O'Rourke and Tuokko (2003a) did a study comparing four abridged versions of the ZBI including the Hébert and Knight models as well as one by Bédard and colleagues (2001) and the aforementioned model by Whitlatch and colleagues (1991). They found that all the models were viable per the confirmatory factor analysis goodness-of-fit indices.

Summary and Hypotheses

There have been several factor analysis studies of the ZBI. Mostly, these analyses have been done to verify or explore a factor structure for the original ZBI or abridged versions in caregiver samples, and other times, they were done to develop briefer or screening versions of the ZBI. Two-factor and three-factor models have been found to fit the data best. There is not yet one model that stands out above the rest and none include all items, but there has been some correspondence across studies regarding factor themes and item content for briefer versions of the ZBI.

The Knight and colleagues model has been replicated more than others in the literature and has fit those data sets moderately well overall. The focus of this paper was to confirm this three-factor model and to test its invariance in Blacks and Whites. Though the need for analysis that confirms the psychometric properties of scales that are used to compare ethnic groups has been noted (see Janevic & Connell, 2001 for review), so far studies have not focused on possible racial/ethnic differences in the factor structure of the 22-item English version of the ZBI. In particular, it has

been suggested that Black and White caregivers may have different perspectives about caregiving (Lawton et al., 1992). This cultural difference in perspective may relate to difference in comprehension and responses to items that might affect the measurement equivalence of the ZBI across groups.

Greater understanding of the factor structure of the ZBI has important implications for interpreting the caregiver literature that uses the ZBI. So far there have not been studies to examine whether or not the ZBI operates in an equivalent manner across race groups, though many of the studies utilizing the ZBI also include cross-race comparisons. In response to minimal amount of factor analysis data and the lack of measurement equivalence data, the current study adds to the knowledge of the psychometric properties of the ZBI by testing its metric equivalence in samples of Black and White dementia caregivers. Such equivalence requires the same pattern of factor loadings and factor correlations along with equivalent magnitude of these relationships across groups (Posner, Stewart, Marín, & Pérez-Stable, 2001). Thus, the hypotheses in this study assumed metric invariance by race and were as follows: (a) The number of factors were equivalent across groups; (b) the pattern of factor loadings were equivalent in regards to item composition, rank order, and magnitude; and (c) the relationships between the factors of each measure were equivalent.

Methods

Participants

The participants were Black and White dementia caregivers obtained by combining data from two samples. First, a convenience sample was collected from 1990 to 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 to 1993, participants ($n = 159$) qualified if they were the primary caregivers of an adult 55 years or older 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. These are the samples on which the Knight and colleagues model was developed and confirmed.

A second probability sample, not used in the Knight and colleagues study, was obtained from the Stress, Ethnicity, and Caregiving study (SECS), which was also conducted at USC ($n = 160$). Data were obtained from 2001 to 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 aged 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 older than the age of 55 years and had abnormal memory problems or was diagnosed with dementia.

In the combined data set, all 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 (Dura & Kiecolt-Glaser, 1990). This combination was logical because 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 (ZBI) administered by the same group of people (USC research team). The combination of the samples increased the racial/ethnic representation of the total data set and allowed for greater power for the multigroup invariance testing.

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 as a result were not included.

Burden.—The 22-item ZBI (Zarit et al., 1980) was used to assess appraisal of burden related to caregiving. The items are scored from 0 to 4, with higher scores indicating higher levels of distress. Total scores range from 0 to a maximum score of 88. Relative to a single global rating of burden, the validity of the ZBI has been estimated at .74 and .66 (Bédard et al., 2001).

The internal consistency of the ZBI has been reported as $\alpha = .92$ (Hébert et al., 2000; Knight et al., 2000). The internal consistency for all 22 items for the Black sample in this study is $\alpha = .92$, and $\alpha = .93$ for the White sample. For the items in the Knight model specifically, the α was .88 for the Black caregivers and .90 for the White caregivers and total scores range from 0 to a maximum score of 56.

Missing Data

Of the 414 caregivers interviewed, 227 White caregivers and 181 Black caregivers had ZBI data. All cases with missing data were deleted. In total, six cases (3%) were deleted from the Black ZBI file ($n = 175$), and two cases (<1%) were deleted from the White ZBI file ($n = 225$). Casewise deletion was used because less than 10% of the sample size had missing data. Selection bias due to missing data was not considered an issue as there was no systematic pattern to the missing data.

Analysis

CFA using AMOS 4.0 was used to examine the Knight 14-item three-factor model. Knight and colleagues (2000) set the two loadings of their Self-Criticism factor equal because there were only two observed variables with a similar scale for that factor. However, setting the loadings equal to one another was not possible with AMOS because it led to those two parameters being unidentified in the analysis. Thus, in the analyses for this study, one of these items was instead set to 1 to set the scale for that factor and to obtain an identified model.

For the first hypothesis, to test that the pattern of items and factors for this scale was equivalent across groups, 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 ZBI. To test the second hypotheses, both caregiver samples were analyzed simultaneously as in the analysis for the first hypothesis, but this time, 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. Because these analyses focus on metric equivalence, 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 CFA 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 & Lomax, 1996). We include each type to compare and assess the fit of the ZBI models to our samples of dementia caregivers.

One index is the χ^2/df ratio. A small value indicates good fit, and a large one indicates poor fit (Byrne, 2001). Conservative standards suggest that a χ^2/df ratio of no more than three serves as an adequate indicator of good fit (Carmines & McIver, 1981; Garson, 2004).

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, and values in the .50s are acceptable (Byrne, 2001).

The CFI measures the comparative reduction in lack of fit by a hypothesized model versus a baseline model (Hoyle & Panter, 1995). Values for models with acceptable fit should be .90 or greater with higher values being recommended (Hu & Bentler, 1999). This index is independent of sample size and takes model complexity into account.

Table 1. Demographic Variables of Caregivers and Dependent Variables by Race

	White (<i>n</i> = 225)		Black (<i>n</i> = 175)			
Background characteristics	<i>SD</i>		<i>SD</i>		Chi-square	<i>t</i> test
Female caregivers (%)	68.9	—	76.0	—	2.47	
Age ^a	62.88	(13.57)	57.48	(14.44)		3.83***
Self-reported burden—ZBI						
22-item	40.51	(16.92)	33.82	(16.43)		3.97***
Knight	25.65	(11.02)	21.32	(10.47)		3.99***

Notes: Categorical variable “female” is reported with percentage and a chi-square test. Other entries reported as means and standard deviations (*SDs*) with *t* tests. ZBI = Zarit Burden Interview.

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

****p* < .001

The root mean square error of approximation (RMSEA), an index of model parsimony, measures the lack of fit per *df* (MacCallum, 1995). Values less than .05 indicate good fit. Values up to .08 indicate reasonable errors of approximation in the population, whereas those above .1 represent poor fit. With some agreement on cut-points, 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 CI 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 *df* equal to the consequent difference in *df*. If the difference in chi-square is found not to be significant, then the hypothesis of an invariant pattern of loadings is considered tenable (Byrne, Shavelson, & Muthén, 1989).

Results

Sample Characteristics

White caregivers were significantly older than Black caregivers, the majority of both groups of caregivers were female. Mean burden scores on the 22-item ZBI were 33.82 for Black caregivers and 40.51 for White caregivers of a maximum of 88. Of a maximum score of 56, mean scores for the Knight and colleagues' 14 items for the Black and White caregivers are 21.32 and 25.65, respectively. The two groups differed significantly on their ZBI total scores, with White caregivers scoring higher, regardless of number of items or model used. Scores between 21 and 40 have been proposed as indicating mild to moderate burden in the 22-item scale (Hébert et al., 2000). Both groups seem to fall into this group, with White caregivers right at the limit. This coincides with the literature that has found dementia caregivers to have high levels of burden. See Table 1 for specific results. The data were somewhat skewed for both groups, but the distributions did not depart substantially from normality.

Factor Analysis Results

Fourteen ZBI items were included in the analysis. The three factors in this model were called

Table 2. Fit Statistics for CFAs of the Zarit Burden Interview Three Factors With Various Levels of Constraint Including Both Samples

Model	χ^2	<i>df</i>	Ratio (χ^2/df)	PGFI	CFI	RMSEA
Model 1a	376.825	148	2.55	0.618	0.903	0.062
Model 2b	391.876	159	2.46	0.661	0.901	0.061
Model 3c	393.931	162	2.43	0.673	0.902	0.060

Notes: *N* = 400. CFI = comparative fit index; PGFI = parsimony goodness-of-fit; RMSEA = root mean square error of approximation. a = no constraints, b = factor loadings (item regression weights) constrained equal, and c = factor loadings and factor covariances constrained.

Table 3. Unstandardized Factor Loadings for the Zarit Burden Interview Constrained Equal Across Black and White Caregiver Groups

Three-factor Model Names and 14 Items		Factor Loadings
Item #	Item name	
Factor: Embarrassment/Anger		
4	Feel embarrassed	1.00
5	Feel angry	1.15
6	Relative affects relationships with others	1.38
9	Feel strained	1.38
10	Health has suffered	1.51
11	Don't have privacy	1.61
13	Feel uncomfortable having friends over	1.33
18	Wish could leave care to someone else	1.34
Factor: Patient's Dependency		
2	Don't have time for self	1.00
8	Relative is dependent	0.45
12	Social life has suffered	1.10
14	Only one relative depends on	0.87
Factor: Self-Criticism		
20	Should be doing more	1.00
21	Could do a better job	0.86

Notes: Only one set of coefficients is reported because parameters for Black group and White group are equal.

Embarrassment/Anger, Patient's Dependency, and Self-Criticism. They were composed of eight, four and two items, respectively (See Table 3).

Hypothesis 1.—To test the first hypothesis, both the Black and the White caregiver samples were analyzed simultaneously for model adequacy. The model was viable as all the items and error estimates differed significantly from zero per the critical ratio (CR) values. The covariance matrices were positive definite for the Black sample. The estimation process was complete for both groups of caregivers.

The covariances among the three factors were somewhat larger for the White group than for the Black one. For both groups, the strongest relationship was between the two factors Embarrassment/Anger and Patient's Dependency. The weakest covariance for both samples was between Patient's Dependency and Self-Criticism.

The $\chi^2/df = 2.55$, the PGFI was .618, and the CFI was .903. All three supported model fit. The RMSEA was good at .062 and fell within a CI that ranged from .055 to .070. This RMSEA can be reported with 90% confidence that the true value falls within those bounds, which represents a good degree of precision. Altogether, the model was adequate. Given these findings, the first hypothesis

was supported, with the number of factors for the Knight model being equivalent across groups.

Hypothesis 2.—The Black and the White caregiver samples were analyzed simultaneously as before, but this time, the factor loadings also were constrained across the two groups. Generally, the parameters for both samples were normal. The estimates were significant (per CR), and their direction and magnitude appeared satisfactory. The *df* increased by 11 when the item regression weights were constrained equal across groups (11 items were constrained equal and three items remained fixed at 1 to set factor scales). All matrices were positive definite for both groups. The correlation between Embarrassment/Anger and Patient's Dependency was again the highest among the factor correlations for both the Black and the White caregivers (.835 and .901, respectively).

The $\chi^2/df = 2.46$, the PGFI was .661, and the CFI was .901. The RMSEA remained almost the same as in the previous analysis at .061. Taken together, the values of the indices denoted that the model did fit the data and that the Knight model was equivalent in terms of factor composition and loadings across groups. Table 2 shows the comparison between the model tested with no constraints and this one. Table 3 shows which items load onto each factor and the factor loadings when constrained equal across the two groups. This model's χ^2 was not significantly greater than that for the baseline model. This finding supported invariance of the item factor loadings in the Knight model.

Hypothesis 3.—For the third analysis of the Knight model, the factor loadings and the covariances between the three factors were constrained as equal across the two groups. Overall, the parameters for both samples were reasonable. They were all significant and in the right direction.

The *df* increased by 3 when the factor covariances were constrained equal across groups. The $\chi^2/df = 2.43$, the PGFI was .673, and the CFI was .902. Again, the PGFI and the CFI supported model fit. The RMSEA was almost the same as in the previous analyses at .060. Altogether, the indices were evidence for adequate model fit. Thus, the relationships between the three factors were also found to be equivalent for the two groups for the Knight model. See Table 4 for the factor covariances constrained equal across the two groups. Again, there is evidence of invariance for this model.

Table 4. Factor Covariances for Zarit Burden Interview When Factor Loadings and Covariances Are Constrained Equal Between Black And White Caregivers

	Embarrassment/ anger	Patient's dependency	Self- criticism
Embarrassment/ anger	—	.50	.24
Patient's dependency	—	—	.23
Self-criticism	—	—	—

Summary of Findings for Hypotheses.—In general, the models were adequate. The PGFIs were consistently above .50, the CFIs were at .90, and the RMSEA values indicated good fit across all three analyses. The unconstrained as well as the two constrained analyses all had complete estimation and parameters within normal range and in the right direction. The models for Hypotheses 2 and 3 did show improved fit via added constraints. The chi-square statistics increased in these analyses, but the trade-offs for parsimony were good ones because the increases in chi-square were not significant for both of the constrained models as compared with the baseline model. See Tables 2 and 5 for model comparisons of fit statistics and differences in chi-square at each level of constraint.

Discussion

A measure with cross-racial compatible validity should demonstrate like-factor structures regarding number of factors and similar factor loadings for comparative racial groups (Tran, 1997). The Knight three-factor model has done this by showing in this analysis fit for both Black and White samples of caregivers when factors, factor loadings, and factor covariances were constrained to be equal across both groups. A three-factor model for the ZBI was found not only to fit a diverse caregiver sample but also to be metrically invariant across two race groups not just in the overall placement of the items but in the relationship of items and factors to each other. This study expands upon the invariance literature by submitting the first results of an examination of metric invariance for the ZBI with Black and White dementia caregivers.

These analyses would suggest that the 14 items in the Knight model can be used as a brief reliable measure of burden and may be especially useful in studies with cross-race comparisons. Alternatively,

Table 5. Chi-Square Difference Tests for the Zarit Burden Interview Three-Factor Models At Various Levels of Constraint

Model	χ^2	df	p value
Model 1a			Baseline model
Model 2b	15.05	11	>.05
Model 3c	17.11	14	>.05

Notes: a = no constraints; b = factor loadings (item regression weights) constrained equal; c = factor loadings and factor covariances constrained.

for studies where the 22-item version is preferred, the results obtained from the 22-item version can be compared with the invariant 14-item version to see if the 14-item and 22-item versions have similar results. This would give further credence to the 22-item results. For example, in this study, the mean total scores for both versions showed that both Black and White caregivers were mildly to moderately burdened, and their scores were significantly different for both versions with White caregivers being more burdened.

It has also been noted that there have been attempts to establish an abbreviated burden measures for screening purposes as some have suggested that the original ZBI may take too long to administer when time with caregivers is limited (Bédard et al., 2001; O'Rourke & Tuokko, 2003b). Therefore, these findings also have applied-setting implications such as offering a reliable, valid, and brief version of the ZBI to use with clients especially when time is scarce, for example, during care recipient doctor visits and case worker appointments. A briefer version of the ZBI would also be an efficient tool for shorter protocols in research studies.

So far, all the published factor structures for ZBI have been for abridged versions. Because the 22-item version of the ZBI is common in the literature (Knight et al., 2000) and has been shown to be reliable across various populations of caregivers and patients (Bachner & O'Rourke, 2007), it would be useful to have a factor structure for all the ZBI items minus the global item. Future research is needed to see if the 21 items will fall into a stable pattern with multiple samples and then if such a model also showed invariance across various groups.

Practically, having an accepted factor structure for the ZBI across race groups is important because knowing which dimensions of burden contribute more or less to a total score provides more detailed

information to a researcher or any other professional who works with caregivers about the impact of caregiving on the caregiver (Novak & Guest, 1989). Such data may allow for increased ability to parse out patterns of caregiving effects or correlates and antecedents of caregiver burden (George & Gwyther, 1986).

As noted earlier, most caregiver studies that used the ZBI and included cross-race comparisons found that White caregivers were higher on burden. A few found that the two groups were similar, but never did Black caregivers have higher levels of burden as measured by the ZBI. This study examined the factor structure of the ZBI to contribute to the discussion about whether or not inconsistent findings may be related to inconsistent measurement of burden across these two groups. These results provide support for the validity of the findings where burden was lower for Blacks.

Limitations

The sample of Black caregivers was less than 200. Though some fit indices 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. Reporting a variety of fit indices was an effort to address this matter. As well, because CFA operates within the framework of structural equation modeling, which is based on asymptotic (large sample) rules, when sample size is smaller, estimators may not operate as expected (Byrne, 2001). The data were ordinal and slightly skewed. However, the ZBI had more than four categories and skewness < 1, the conditions under which estimation processes are least affected (Byrne, 2001). ML estimation method has been shown to be robust to skewed data (Chou & Bentler, 1995).

Lastly, the study sample was regional to the western United States, with most participants living in the Los Angeles County area. Almost 40% 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. Furthermore, because the sample in this study was of dementia caregivers, caution should be taken in applying the findings of this report to other types of caregivers.

Implications

Studies such as the present one show that the ZBI, though often used as a single score scale, likely has several dimensions of burden. Furthermore, such factors have promise for effective use not just with White groups but with other races and ethnicities as well. Future research should consider replicating and testing this three-factor, briefer version of the ZBI for invariance in other groups.

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 ZBI is very capable of measuring symptoms of burden in both types of caregivers.

Regarding policy, these findings are also helpful information. In order to know prevalence rates and devise treatment plans and services for caregivers, it is important that scales measure the constructs correctly and the same way for various people. If not, then rates of distress could very well be biased or misestimated for those groups (Mui et al., 2001). Erroneous conclusions drawn from measures with dissimilar properties across groups can have wide-ranging implications for public health policy (Posner et al., 2001). Accurate detection gives policymakers the information they need to make decisions regarding the resources for which they will advocate on behalf of the caregivers in their constituency.

Funding

This work was supported by grants from the National Institutes of Health: African-American and White Caregivers: Health Effects (R01 AG 016307-01), Multidisciplinary Research Training in Gerontology (T32 AG 000037-26), and Interdisciplinary Women's Health Research at Medical University of South Carolina (K12 HD055885-03).

References

- Allen-Kelsey, G. J. (1998). Caregiver burden among African-American and Anglo-American family caregivers. *Topics in Geriatric Rehabilitation, 14*, 63-73.
- Bachner, Y. G., & Ayalon, L. (2010). Initial examination of the psychometric properties of the short Hebrew version of the Zarit Burden Interview. *Aging and Mental Health, 14*, 25-30. doi: 10.1080/13607861003601841
- Bachner, Y. G., & O'Rourke, N. (2007). Reliability generalization of responses by care providers to the Zarit Burden Interview. *Aging & Mental Health, 11*, 678-685. doi:10.1080/13607860701529965
- Bédard, M., Molloy, D. W., Squire, L., Dubois, S., Lever, J. A., & O'Donnell, M. (2001). Zarit Burden Interview: A new short version and screening version. *The Gerontologist, 41*, 652-657.

- Bell, C. M., Araki, S. S., & Neumann, P. J. (2001). The association between caregiver burden and caregiver health-related quality of life in Alzheimer disease. *Alzheimer Disease and Associated Disorders*, 15, 129–136.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456–466.
- Carmine, E., & McIver, J. (1981). Analyzing models with unobservable variables. In G. W. Bohrnstedt, & E. F. Borgotta (Eds.), *Social measurement: Current issues* (pp. 65–115). Beverly Hills, CA: Sage.
- Chou, C-H, & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37–55). Thousand Oaks, CA: Sage.
- Dura, J. R., & Kiecolt-Glaser, J. K. (1990). Sample bias in caregiving research. *Journal of Gerontology*, 45, P200–P204.
- Fredman, L., Daly, M. P., & Lazur, A. M. (1995). Burden among White and Black caregivers to elderly adults. *Journal of Gerontology*, 50B, S110–S118.
- Garson, G. D. (2004). *Structural equation modeling example using WinAMOS (webpage)*. Retrieved from <http://www2.chass.ncsu.edu/garson/pa765/semAMOS1.htm>
- George, L. K., & Gwyther, L. P. (1986). Caregiver well-being: A multidimensional examination of family caregivers of demented adults. *The Gerontologist*, 26, 253–259.
- Hébert, R., Bravo, G., & Prévaille, M. (2000). Reliability, validity and reference values of the Zarit Burden Interview for assessing informal caregivers of community-dwelling older persons with dementia. *Canadian Journal on Aging*, 19, 494–507.
- Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.
- Hu, L-T, & Bentler, P. M. (1999). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.
- Institute of Medicine. (2007). *The Future of Disability in America*. Washington, DC: The National Academies Press.
- Janevic, M. R., & Connell, C. M. (2001). Racial, ethnic, and cultural differences in the dementia caregiving experience: Recent findings. *The Gerontologist*, 41, 334–347.
- Knight, B. G., Fox, L. S., & Chou, C-P (2000). Factor structure of the Burden Interview. *Journal of Clinical Geropsychology*, 6, 249–258.
- Kosberg, J. I., Kaufman, A. V., Burgio, L. D., Leeper, J. D., & Sun, F. (2007). Family caregiving to those with dementia in rural Alabama: Racial similarities and differences. *Journal of Aging Health*, 19, 3–21.
- Kumamoto, K., & Arai, Y. (2004). Validation of 'personal strain' and 'role strain': Subscales of the short version of the Japanese version of the Zarit Burden Interview (J-ZBI_8). *Psychiatry and Clinical Neurosciences*, 58, 606–610. doi:10.1111/j.1440-1819.2004.01310.x
- Lawton, M. P., Kleban, M. H., Moss, M., Rovine, M., & Glicksman, A. (1989). Measuring caregiving appraisal. *Journal of Gerontology: Psychological Sciences*, 44, P61–P71.
- Lawton, M. P., Rajagopal, D., Brody, E., & Kleban, M. H. (1992). The dynamics of caregiving for a demented elder among black and white families. *Journal of Gerontology: Social Sciences*, 47, S156–S164.
- MacCallum, R. C. (1995). Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 16–36). Thousand Oaks, CA: Sage.
- Miller, E. A., Rosenheck, R. A., & Schneider, L. S. (2010). Caregiver burden, health utilities, and institutional service costs among community-dwelling patients with Alzheimer disease. *Alzheimer Disease and Associated Disorders*, 24, 380–389. doi:10.1097/WAD.0b013e3181eb2f2e
- Mohamed, S., Rosenheck, R., Lyketsos, C. G., & Schneider, L. S. (2010). Caregiver burden in Alzheimer disease: Cross-sectional and longitudinal patient correlates. *American Journal of Geriatric Psychiatry*, 18, 917–927.
- Mui, A. C., Burnette, D., & Chen, L. M. (2001). Cross-cultural assessment of geriatric depression: A review of the CES-D and the GDS. *Journal of Mental Health and Aging*, 7, 137–164.
- Novak, M., & Guest, C. (1989). Application of a multidimensional Caregiver Burden Inventory. *The Gerontologist*, 29, 798–803.
- O'Rourke, N., & Tuokko, H. A. (2003a). Psychometric properties of an abridged version of the Zarit Burden Interview within a representative Canadian caregiver sample. *The Gerontologist*, 43, 121–127.
- O'Rourke, N., & Tuokko, H. A. (2003b). The relative utility of four abridged versions of the Zarit Burden Interview. *Journal of Mental Health and Aging*, 9, 55–64.
- Pinquart, M., & Sorensen, S. (2005). Ethnic differences in stressors, resources, and psychological outcomes of family caregiving: A meta-analysis. *The Gerontologist*, 45, 90–106.
- Posner, S. F., Stewart, A. L., Marín, G., & Pérez-Stable, E. J. (2001). Factor validity of the Center for Epidemiological Studies Depression Scale (CES-D) among Urban Latinos. *Ethnicity and Health*, 6, 137–144.
- Schulz, R., O'Brien, A. T., Bookwala, J., & Fleissner, K. (1995). Psychiatric and physical morbidity effects of dementia caregiving: Prevalence, correlates, and causes. *The Gerontologist*, 35, 771–791.
- Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Tran, T. V. (1997). 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*, 7, 500–517.
- Whitlatch, C. J., Zarit, S. H., & von Eye, A. (1991). Efficacy of interventions with caregivers: A reanalysis. *The Gerontologist*, 31, 9–14.
- Zarit, S. H., Reever, K., & Bach-Peterson, J. (1980). Relatives of impaired elderly: Correlates of feelings of burden. *The Gerontologist*, 20, 373–377.