Medicine®

OPEN

Religiosity in black and white older Americans
Measure adaptation, psychometric validation, and racial difference

Chengwu Yang, MD, MS, PhD[a], Marvella E. Ford, PhD[b], Barbara C. Tilley, PhD[c], Ruth L. Greene, EdD[d*]

Abstract

Racial difference of religiosity in a heterogeneous older population had long been a focal point of gerontological research. However, most religiosity measures were developed from homogenous sample, few underwent rigorous psychometric validation, and studies on racial difference of religiosity had been obstructed. This cross-sectional study adapted a religiosity measure originally designed for blacks only to a heterogeneous older population of blacks and whites, validated its psychometric properties, and examined racial difference of religiosity. Based on qualitative research of concepts, intensive literature review, and abundant experiences in this field, we adapted the original measure. Then, using the data collected from a survey of 196 black and white Americans 55 years and older in Charlotte, North Carolina, we investigated full-scale psychometric properties of the adapted measure at the item-, domain-, and measure- level. These psychometric validations included item analysis, item-scale correlations, correlation matrix among items, confirmatory factor analysis (CFA) to determine if the original factor structure held after adaptation, and reliability analysis using Cronbach's alpha. Finally, using *M*ultiple *I*ndicators and *M*ult*i*ple *C*auses (MIMIC) models, we examined racial difference of religiosity through regression with latent variables, while potential measurement bias by race through differential item functioning (DIF) was adjusted in the MIMIC models. In result, we successfully adapted the original 12-item religiosity measure for blacks into an 8-item version for blacks and whites. Although sacrificed few reliability for brevity, the adapted measure demonstrated sound psychometric properties, and retained the original factor structure. We also found racial differences of religiosity in all three domains of the measure, even after adjustment of the detected measurement biases in two domains. In conclusion, the original measure can be adapted to and validated for a heterogeneous older population of blacks and whites. Although the adapted measure can be used to measure the three domains of religiosity in blacks and whites, the observed racial differences of religiosity need to be adjusted for measurement biases before meaningful comparisons.

Abbreviations: CFA = confirmatory factor analysis, DIF = differential item functioning, MIMIC = *M*ultiple *I*ndicators and *M*ult*i*ple *C*auses, NOR = nonorganizational religiosity, NSBA = National Survey of Black Americans, OR = organizational religiosity, SR = subjective religiosity.

Keywords: African Americans, aging, differential item functioning (DIF), factor analysis, measurement bias, MIMIC models, religiosity

[a] Department of Public Health Sciences & Office for Scholarship in Learning and Education Research, College of Medicine, The Pennsylvania State University, Hershey, PA, [b] Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, [c] Division of Biostatistics, School of Public Health, University of Texas Health Science, Center at Houston, Houston, TX, [d] Department of Psychology, Johnson C. Smith University, Charlotte, NC.

* Correspondence: Ruth L. Greene, Johnson C. Smith University, 100 Beatties Ford Road, Charlotte, NC 28261 (e-mail: rgreene@jcsu.edu).

1. Introduction

Measurement is of vital importance for science,[1] including study of religion. Since most measures were developed from homogeneous samples, their validity in diverse populations is a big concern.[2–4] While measurement in diverse populations is still an emerging science with research on measure adaptation clearly at its cutting-edge, reporting of this kind of work is missing.[4] On the other hand, the relationship among different aspects of religion to the health status of older adults has long been a focal point in gerontological research and other field such as oncology.[5–9] And some measurement problems of religiosity have been identified.[10]

There has been a long history at the measure of religiosity.[5,11–13] Using a subset (n = 947) of a sample from Kauffman and Harder's 1975 survey of 5 Mennonite and Brethren in Christ denominations, Ainlay and Smith[14] established an 11-item, 3-dimensional measure for religiosity. The 3 domains they established were public participation, attitudes toward participation, and private participation. Realizing the lack of external validity of Ainlay and Smith's measure due to the use of an exclusively Mennonite and Brethren sample, Chatters et al[15,16] proposed and tested another measure using a subset (n = 446) of a sample from the National Survey of Black Americans (NSBA). The Chatters et al's[16] measure had

3 domains that were similar to Ainlay and Smith's: organizational religiosity (OR), nonorganizational religiosity (NOR), and subjective religiosity (SR), but it had 12 items. Using the new measure, Chatters and coworkers[17–19] found that religiosity plays an important role in the protection and maintenance of health among blacks. Although there had been some other measures for religiosity,[20,21] Chatters et al's[16] measure had been very influential in this field. However, a limitation of the Chatters et al's measure is its use in an exclusively black sample. Therefore, its validity in a broader and diverse population is of concern, but this topic has not been studied.

Basically there are 3 options when applying a race-specific measure to diverse populations: use it "as is," develop a new one, or adapt it. Because the first option has obvious risks and limitations and the second one is usually too challenging, the third option is most realistic.[4] Also, as demonstrated by the whole history of the Chatters and coworkers[15,16,19,22] measure, adaptations and modifications had been constantly implemented.

Based on a cross-sectional design, this study adapts the Chatters et al's[16] black-specific religiosity measure for use in a sample of black and white older Americans, with full-scale psychometric assessments at the item-, domain-, and measure-level. Also, using the adapted measure, this study examines racial differences in religiosity while adjusting for measurement bias.

We hypothesis that the original measure can be adapted, that the adapted one will have sound psychometric properties, and that racial difference among blacks and whites will be detected by the adapted measure.

2. Methods

2.1. Measure adaptation

Given this study has a diverse black and white sample while the original measure was developed from an exclusive black sample, adaptation of the measure is obviously necessary. However, the adaptations in this study were rigorous, based on qualitative research of concepts, intensive literature review, and authors' abundant experiences in this field.[4]

In the Chatters et al's[16] measure (Table 1), the organizational religiosity (OR) domain included 5 items. Three of the 5 items required respondents to list the number of church clubs they belonged to (Y3), the number of other church activities that they engaged in besides the regular church service (Y4), and whether they held offices in the church (Y5). These 3 items were excluded in the modified scale because several gerontological studies have shown that these items in organizational religious involvement in older adults may be more associated with declining functional

Table 1			
Adaptations of Chatters et al's[16] religiosity measure.			
Dimension	Item	Original variable name and coding	New variable name and coding
Organizational religiosity (OR)	How often do you usually attend religious services?	Y1: 1 = never, 2 = less than once a year, 3 = a few times a year, 4 = a few times a month −1 to 3 times, 5 = at least once a week −1 to 3 times, 6 = nearly every day −4 or more times a week	OR1: 1 = everyday, 2 = weekly, 3 = monthly, 4 = yearly, 5 = less than yearly
	Are you an official member of a church or other place of worship?	Y2: 0 = no, 1 = yes	OR2: 0 = no, 1 = yes
	How many church clubs or organizations do you belong to or participate in?	Y3: the exact number is coded	Excluded
	Besides regular service, how often do you take part in other activities at your place of worship?	Y4: 1 = never, 2 = a few times a year; 3 = a few times a month −1 to 3 times; 4 = at least once a week −1 to 3 times; 5 = nearly every day −4 or more times a week	Excluded
	Do you hold any positions or offices in your church or place of worship?	Y5: 0 = no, 1 = yes	Excluded
Nonorganizational religiosity (NOR)	How often do you read religious books or other religious materials?	Y6: 1 = never, 2 = a few times a year; 3 = a few times a month −1 to 3 times; 4 = at least once a week −1 to 3 times; 5 = nearly every day −4 or more times a week	NOR1: 1 = everyday, 2 = weekly, 3 = monthly, 4 = yearly, 5 = never
	How often do you watch or listen to religious programs on TV or radio?	Y7: same as Y6	NOR2: same as NOR1
	How often do you pray?	Y8: same as Y6	NOR4: same as NOR1
	How often do you ask someone to pray for you?	Y9: same as Y6	NOR3: same as NOR1
Subjective religiosity (SR)	How religious would you say you are?	Y10: 1 = not religious at all, 2 = not too religious, 3 = fairly religious, 4 = very religious	SR2: 1 = very religious, 2 = fairly religious, 3 = not too religious, 4 = not religious
	How important was religion in your home when you were growing up?	Y11: 1 = not important at all, 2 = not too important, 3 = fairly important, 4 = very important	SR1: 1 = very important, 2 = fairly important, 3 = not too important, 4 = not important
	How important is it for Black parents to send or take their children to religious services?	Y12: same as Y11	Excluded

health than in the strength of organizational affiliation.[23–25] Those authors suggested that nonorganizational and SR may increase to offset the decline in disengagement in formal organizational activities. Religious activities could also take on different forms in different cultures, with blacks and whites having different belief systems and domains of religious expression.[22] Therefore, the SR question intended to assess the importance of black parents to send or take their children to religious services (Y12) was excluded.

There had been other considerations to shorten the Chatters et al's[16] measure, which support the elimination of the 4 items (Y3, Y4, Y5, Y12, see Table 1). First, religiosity was only 1 out of many of interest in our original study of older adults, as is often the case in medical research projects. In these contexts, the time that can be allocated to the collection of religiosity data is usually very limited. Second, shorter measures with fewer items reduce the burden on participants,[26] which is especially important when the population includes older adults and those with poor health status. Third, we believe that some of the items in the Chatters et al's[16] measure are either unnecessarily redundant or inappropriate for nonblack participants, and redundancy is not always preferred in measure development.[27]

We also revised the options for 7 of the 8 remaining questions, in order to make the response options more concise and easier to read. This kind of practice, that is, adapting items from existing measures and making appropriate changes in order to better fit the survey populations, is highly recommended and widely implemented.[4,28]

2.2. Field survey and data collection

The protocol of this study was reviewed and approved by the Institutional Review Board (IRB) at the corresponding author's institute before the initiation of field survey. The sample consisted of adults 55 years and older living in Charlotte, NC. It was drawn according to a multistage, area probability procedure designed to insure that every household in Charlotte had the same probability of selection. A primary area was selected for interviewing based on the census distribution of the population. In contrast to the original study by Chatters of only black households, these sites were stratified according to racial composition to include white and black households, and smaller geographical areas "clusters" were randomly chosen. Inhabitable households were listed. Finally, within each inhabitable household, trained interviewers interviewed all household members who were 55 and older. The participants completed an extensive 1-hour questionnaire based on a survey developed at the University of Michigan for the NSBA, which included the just adapted religiosity measure.

2.3. Psychometric validation

The adapted measure was administered to all of the participants, and some basic demographic characteristics were also collected and analyzed. According to recent literatures,[4,29] the modifications to this measure were at a "substantial" level and were made to its contents, contexts, and formats. From the findings of a recent systematic review on spirituality measures,[30] even changes in the measures' content after cultural adaptation were poorly reported, not to mention the adequate assessment of psychometric properties of the adapted measures and the well-known culture bias of spirituality measures. And as Chatters et al[22] pointed out over a decade ago, rigorous and systematic research on religion measures across diverse populations should

be critical for this field. Therefore, a full-scale psychometric assessment of the adapted measure was implemented on the data at the item-, domain-, and measure-level. These included item analysis, item-scale correlations, confirmatory factor analysis (CFA), reliability analysis using Cronbach alpha, and investigation of measurement bias by race through differential item functioning (DIF) analysis using *M*ultiple *I*ndicators and *M*ult*I*ple *C*auses (MIMIC) models. Moreover, racial differences in religiosity were also examined while adjusting measurement bias through the MIMIC models.

2.3.1. Item analysis. The basic psychometric property for an item is its variability.[26,31] For each of the 8 items, frequency table of responses is reported. Optimally, items should not have any of its response categories chosen by either less than 5% or greater than 95% participants.[32]

2.3.2. Item-scale correlations. A good item shall substantially correlate with the domain it belongs to, and item-scale correlation is a measure for this property.[31] There are 2 types of item-scale correlations: uncorrected and corrected. While the uncorrected item-scale correlation represents how representative the item is to the whole scale, the corrected one represents how closely the item is correlated to other items in the same domain.

2.3.3. Correlation matrix among items. Correlations among items play a key role for a measure.[31] A good item should be sufficiently correlated with others, especially with items in the same domain.

2.3.4. Confirmatory factor analysis (CFA). CFA was used to determine if the same factor structure of the Chatters et al's[16] original measure holds after adaptation. The following model fit indexes and the cut-off values for adequate fit were used[33–35]: comparative fit index (CFI) ≥ 0.95, Tucker–Lewis index (TLI) ≥ 0.95, weighted root mean square residual (WRMR) ≤ 1.00, root mean square error of approximation (RMSEA) ≤ 0.08, and normed Chi-square (NC, which equals to the Chi-square divided by the degree of freedom of the model) ≤ 5.0.

2.3.5. Reliability analysis. For each of the domains and the whole measure, Cronbach alpha was used to assess the internal reliability. A Cronbach alpha value around 0.7, 0.8, or 0.9 is considered as an indicator of "adequate," "very good," or "excellent" internal consistency reliability, respectively.[35]

2.4. Racial difference

2.4.1. MIMIC modeling. The MIMIC model can be interpreted as a CFA model with covariates, and it has 2 parts[36–39]: a measurement part and a structural part. The measurement part includes the CFA model and possible direct effects of covariates to the indicators (items) after mediated by the latent traits (i.e., DIF effects).[34] The structural part includes regressions of the latent traits to the covariates and correlations among the latent traits (Fig. 1). While a significant effect of covariate on the latent trait that stands for the population heterogeneity (e.g., racial differences in religiosity) is usually preferred, a significant direct effect from covariate to the items is not, since it stands for measurement bias, that is, DIF. An item (question) shows DIF when the probability of choosing a specific response differed significantly across the groups of interest conditional on the latent trait that the item is measuring.[40,41] In an iterative fashion,[38] DIF analysis was done by checking the modification indices (MI) for each of the 8 possible DIF effects (1 covariate, i.e., race, on
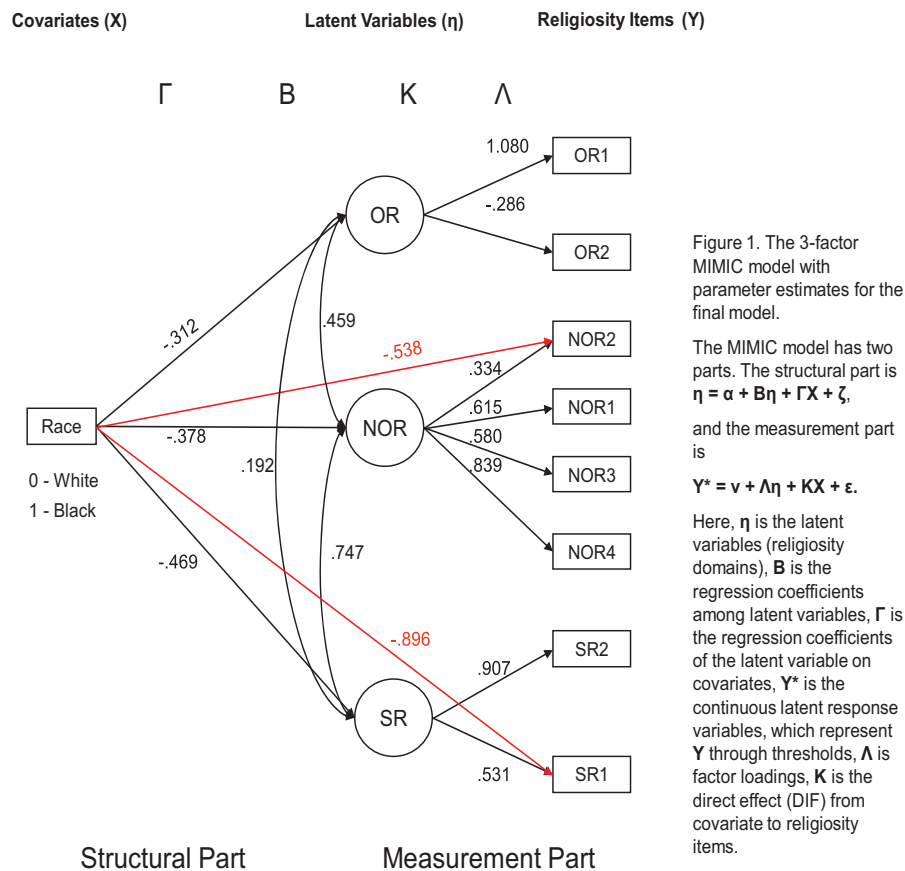
Figure 1. The 3-factor MIMIC model with parameter estimates for the final model.

The MIMIC model has two parts. The structural part is $\eta = \alpha + B\eta + \Gamma X + \zeta$, and the measurement part is $Y^* = \nu + \Lambda\eta + KX + \varepsilon$.

Here, η is the latent variables (religiosity domains), B is the regression coefficients among latent variables, Γ is the regression coefficients of the latent variable on covariates, Y^* is the continuous latent response variables, which represent Y through thresholds, Λ is the factor loadings, K is the direct effect (DIF) from covariate to religiosity items.

Figure 1. The MIMIC model of the adapted religiosity instrument. The 3-factor MIMIC model with parameter estimates for the final model. The MIMIC model has 2 parts. The structural part is $\eta = \alpha + B\eta + \Gamma X + \zeta$, and the measurement part is $Y^* = \nu + \Lambda\eta + KX + \varepsilon$. Here, η is the latent variables (religiosity domains), B is the regression coefficients among latent variables, Γ is the regression coefficients of the latent variable on covariates, Y^* is the continuous latent response variables, which represent Y through thresholds, Λ is the factor loadings, K is the direct effect (DIF) from covariate to religiosity items.

8 items).[37] The MI pertaining to 1 possible DIF effect was calculated by comparing the differences in Chi-squares of the model before and after adding the DIF with a Chi-square distribution at 1 degree of freedom.[37] An MI greater than 3.84 indicates there is a DIF effect. To start, the DIF effect (i.e., a direct effect from race to 1 of the 8 items) with the biggest MI greater than 3.84 was added into the MIMIC model. Then, the remaining MIs were rechecked. This procedure was done iteratively, and was stopped when no more remaining MI was greater than 3.84 (Appendix http://links.lww.com/MD/B193). No MI greater than 3.84 suggests that no DIF was observed. If some DIF are identified, then the relative change of regression coefficients from race to the 3 religiosity domains before and after adding the DIF effects to the model will be used to assess the degree to which the population differences in the religiosity domains are contaminated by DIF.[38]

SAS (version 9.3) was used for data preparation, descriptive analyses, and reliability analyses. M*plus* (version 7.1) was used for CFA and MIMIC modeling. For ease of CFA and MIMIC modeling, all of the observations that had at least 1 of the 8 measure items unanswered or had missing values on race were eliminated. Given the relatively small sample size, in order to avoid susceptible influence of outliers, Mahalanobis distance[35] based on the scores of items in the adapted measure was used to detect and exclude multivariate outliers. Basic demographic characteristics of the excluded subjects are reported in comparison with those of included subjects.

3. Results

3.1. Measure adaptation

In summary, the adaptations to Chatters et al's[16] religiosity measure (Table 1) included elimination of 4 (Y3, Y4, Y5, Y12) of the 12 original items, and revision of the options for 7 of the 8 remaining items. These resulted in an 8-item measure of religiosity for its 3 domains. Details are listed in Table 1.

3.2. Demographic characteristics of the study sample

Eighty-three geographic areas in Charlotte were selected for interviewing, and a total of 263 older adults were interviewed. Of the 263, 49 participants had at least 1 of the 8 questions in the religiosity measure unanswered, with the majority did not answer any of the 8; another 8 participants had race unrecorded. Although M*plus* can handle missing values, for quality control purposes, these 57 participants were excluded in psychometric analyses. In addition, 10 other participants were detected as multivariate outliers as indicated by the Mahalanobis distance[35] based on the 8 item scores ($D^2 > 26.09$, df = 8, $P < 0.01$). Therefore, 196 participants remained. The basic demographic characteristics of the 196 included subjects and 67 excluded subjects are reported in Table 2. Overall, there were no statistically significant differences in gender, education, or income between the included and excluded subjects. Although there were some statistically significant difference in race and marital status, the P values were close to 0.05.

3.3. Psychometric properties

3.3.1. Item analysis, item-scale correlation, and correlation matrix.
Some of the psychometric properties of the adapted measure are summarized in Table 3. These include a frequency table of responses for each of the 8 items, uncorrected and corrected item-scale correlations, and the correlation matrix among the 8 items and with the whole measure. Five of the 8 items have each of their response categories fall in the range of 5% to 95%, as does the vast majority (29 out of 35, 83%) of the overall response categories across these 8 items. The uncorrected item-scale correlations range from 0.42 to 0.96, and the corrected item-scale correlations ranged from 0.16 to 0.39. Within each of the 3 domains, items are significantly correlated, and each of the 8 items is significantly correlated with the whole measure ($P < 0.05$).

3.3.2. CFA.
CFA based on the original measurement model of Chatters et al's[16] were conducted on whites (n = 119) and overall sample (n = 196), and the model fit indexes and factor loadings are summarized in Table 4. For blacks, CFA failed due to nonconvergence in the parameter estimation. For whites, the model fit indices are good, with CFI = 0.946, TLI = 0.911, RMSEA = 0.070, WRMR = 0.566, and NC = 1.578. The factor loadings of all of the 8 items are salient and significant, ranged from 0.424 to 0.908. For the overall sample, the model fit indices are excellent, with CFI = 0.978, TLI = 0.963, RMSEA = 0.038, WRMR = 0.516, and NC = 1.283. The factor loadings of all of the 8 items are salient, ranged from 0.231 to 1.100. All of these model fit indices satisfy their corresponding cut-off values stated earlier. These findings indicate that after adaptation, the measure sustained the original 3-dimensional structure for both of the overall sample and the white subset, but the conclusion to black subset is unknown due to the limitation of sample size.

3.3.3. Reliability.
The Cronbach alphas of the 3 domains and the overall measure are also reported in Table 4 for white and overall sample. For the whole measure on white and overall sample, and NOR on white, the Cronbach alphas are close to 0.70, which implied close to "adequate" internal consistency of the corresponding domains and measure. However, other Cronbach alphas indicated less than "adequate" internal consistency for other domains.

3.4. Racial difference

3.4.1. MIMIC modeling and racial differences.
Results for the iterative DIF analysis using MIMIC model in Mplus are reported in the Appendix http://links.lww.com/MD/B193. DIF effects for race were detected for SR1 (How important was religion in your home when you were growing up?) and NOR2 (How often do you watch or listen to religious programs on TV or radio?). At the same level of religiosity, blacks tend to rate religion more important when growing up than whites do (SR1), and blacks tend to watch or listen to religious programs on TV or radio (NOR2) more often than whites do (Fig. 1). The model fit indices and standardized parameter estimates using the variances of the latent trait (religiosity) for the baseline model (without DIF effects) and the final model (with all DIF effects) are reported in Table 5. Standardized parameter estimates are used since the 2 models have a different residual variance for the latent trait (religiosity). The parameter estimates are not standardized based on variances of both the latent trait and the covariates because the covariate (race) is binary.[37] The baseline model does not fit the data well, with CFI = 0.886, TLI = 0.813, RMSEA = 0.079, WRMR = 0.786, and NC = 2.235. This can be interpreted as an indication of possible DIF effects, although the standardized factor loadings are still salient, ranging from 0.277 to 1.113. The addition of 2 DIF effects into the final model improves the model fit substantially, with CFI = 0.976, TLI = 0.956, RMSEA = 0.038, WRMR = 0.549, and NC = 1.289, implying the final model with DIF effects fits the data very well. The factor loadings remain salient with standardized factor loadings ranging from 0.286 to 1.080. The differences in religiosity between the 2 races are represented by the regression coefficients of race on the 3 religiosity domains in the MIMIC model, and the impact of DIF on these differences can be judged by comparing the relative

Table 2

Demographic characteristics of the study sample (N = 263).

Variable	Included (N = 196)		Excluded (N = 67)		P
	N	%	N	%	
Race					
White	119	60.7	37	77.1	0.04[*]
Black	77	39.3	11	22.9	
Gender					
Male	58	29.6	15	31.3	0.82[*]
Female	138	70.4	33	68.7	
Marital status					
Married	105	56.8	40	72.7	0.04[*]
Widow	54	29.2	7	12.7	
Other	26	14.0	8	14.6	
Education					
Mean ± SD	11.3 ± 4.2		11.7 ± 4.2		0.53[†]
Range	0–17		0–17		
N	196		48		
Income					
Mean ± SD	11.7 ± 8.8		13.0 ± 11.2		0.48[†]
Range	0–30		0–30		
N	137		34		

[*] Chi-square test.
[†] t test.

Table 3

Distribution of responses, item-scale correlations, correlation matrix, and Cronbach alpha of the adapted measure among the study sample (N = 196).

Dimension	Item	Distribution of responses (% of each category)					Item-scale correlation: uncorrected	Item-scale correlation: corrected	Correlation matrix							
		1	2	3	4	5			OR1	OR2	NOR1	NOR2	NOR3	NOR4	SR1	SR2
OR	OR1	5	52	22	11	10	0.96	0.16	1.00							
	OR2	11	89	NA	NA	NA	0.42	0.25	0.16*	1.00						
NOR	NOR1	44	26	15	9	6	0.68	0.34	0.32*	0.11	1.00					
	NOR2	23	35	14	12	16	0.68	0.29	0.07	−0.10	0.17*	1.00				
	NOR3	14	20	22	20	24	0.74	0.39	0.25*	0.02	0.32*	0.26*	1.00			
	NOR4	93	4	1	2	0	0.47	0.32	0.23*	0.05	0.27*	0.20*	0.22*	1.00		
SR	SR1	76	17	6	1	NA	0.83	0.34	0.08	0.00	0.14*	0.24*	0.13	0.28*	1.00	
	SR2	49	45	6	0	NA	0.81	0.34	0.17*	0.08	0.27*	0.26*	0.32*	0.32*	0.34*	1.00
Whole measure		NA					NA	NA	0.54*	0.15*	0.66*	0.59*	0.69*	0.50*	0.43*	0.57*

Note: Please refer to Table 1 for the item names and abbreviations of the domain names.
NA = not applicable.
* P < 0.05.

change in these standardized regression coefficients before and after the addition of DIF effects into the MIMIC model. From Table 5, we can find that blacks are more religious than whites at each of the 3 domains (OR, NOR, and SR). (Note: Race is coded as "0—White, 1—Black," and the items are coded in a reverse direction as listed in Table 1.) Moreover, the 2 DIF effects inflated the racial difference in nonorganizational religiosity (NOR) and

in SR, and they almost had no impact on the racial difference in OR (0.312 vs 0.307). Over 33% of the observed racial difference in NOR can be attributed to the effects of DIF [(0.567–0.378)/0.567 = 0.333], and this number in SR is over 40% [(0.788–0.469)/0.788 = 0.405].

4. Discussion

The purposes of this study were to adapt a race-specific religiosity measure in elders for use in diverse Black and White population with full-scale psychometric assessment, and to appropriately examine racial differences in religiosity among older Americans while controlling for measurement bias. The one-third decrease in the number of items from 12 to 8 makes the adapted measure more efficient for administration in medical research, where measures can become lengthy due to the addition of other items reflecting other domains of interest. And the extensive full-scale psychometric assessment of the adapted measure demonstrated that it was valid for religiosity in the study sample. Further, the results show that racial differences are evident in all of the 3 religiosity domains although substantial measurement bias (DIF) is contaminating the differences in 2 domains. Therefore, cautions should be warranted when examining racial differences in scores from the adapted measure, because measurement biases (DIF) were identified in 2 of the 8 items (NOR2 and SR1), which substantially contaminated the racial difference in 2 of the 3 domains: NOR and SR.

Consistent with earlier studies, this study found that blacks are more religious than whites. For example, from a nationwide survey of 748 older whites and 752 older African Americans, Krause[42] found that blacks were more involved in church participation than whites. And, based on data of 1439 elders from the National Survey of American Life (NSAL), Taylor et al[43] also found that older African Americans reported higher levels of religious participation, coping, and spirituality than older whites. Moreover, also from the NSAL, Taylor and Chatters[44] found that blacks watched religious television programs and listened to religious radio programs significantly more frequently than whites. At regional level, Arcury et al[6] found in their sample of 701 community-dwelling elders with diabetes in 2 rural North Carolina counties, African Americans had higher levels of private religious practice than was reported by Native Americans or Caucasians, and Native Americans

Table 4

CFA results and Cronbach alphas for the adapted religiosity measure (196 participants, 8 items, 3 domains).

	White only (N = 119)	Black and White (N = 196)
CFA		
Model fit indexes		
CFI	0.946	0.978
TLI	0.911	0.963
RMSEA	0.070	0.038
WRMR	0.566	0.516
Chi-square	26.827	21.812
df	17	17
NC = Chi-square/df	1.578	1.283
Factor loadings (standard error)		
OR		
OR1	0.869 (0.287)	1.100 (0.789)
OR2	−0.424 (0.196)	−0.231 (0.212)
NOR		
NOR1	0.684 (0.068)	0.577 (0.063)
NOR2	0.468 (0.081)	0.404 (0.071)
NOR3	0.530 (0.076)	0.574 (0.060)
NOR4	0.908 (0.158)	0.854 (0.138)
SR		
SR1	0.907 (0.119)	0.830 (0.088)
SR2	0.512 (0.110)	0.580 (0.094)
Cronbach alpha		
OR	0.397	0.281
NOR	0.621	0.560
SR	0.524	0.504
Whole instrument	0.681	0.646

Note: The CFA model for black only cannot be estimated due to small sample size.
Please refer to Table 1 for the item names and abbreviations of the domain names.
CFA = confirmatory factor analysis, CFI = comparative fit index, NC = normed Chi-square, RMSEA = root mean square error of approximation, TLI = Tucker–Lewis index, WRMR = weighted root mean square residual.

Table 5

Baseline and final fitted MIMIC model results (N = 196).

Model parameter	Baseline model estimate	Final model estimate
Model fit indexes		
CFI	0.886	0.976
TLI	0.813	0.956
RMSEA	0.079	0.038
WRMR	0.786	0.549
Chi-square	49.165	25.772
Df	22	20
NC = Chi-square/df	2.235	1.289
Measurement component		
Factor loadings		
OR		
OR1	1.113	1.080
OR2	−0.277	−0.286
NOR		
NOR1	0.616	0.615
NOR2	0.378	0.334
NOR3	0.582	0.580
NOR4	0.855	0.839
SR		
SR1	0.633	0.531
SR2	0.848	0.907
Direct effects of race (DIF)		
SR1	NA	−0.896
NOR2	NA	−0.538
Structural component		
Race (reference group: white)		
OR	−0.307	−0.312
NOR	−0.567	−0.378
SR	−0.788	−0.469
Correlation among dimensions		
OR with NOR	0.441	0.459
OR with SR	0.195	0.192
NOR with SR	0.802	0.747

Note: Please refer to Table 1 for the item names and abbreviations of the domain names.
CFI = comparative fit index, MIMIC = Multiple Indicators and Multiple Causes, NA = not applicable, NC = normed Chi-square, RMSEA = root mean square error of approximation, TLI = Tucker–Lewis index, WRMR = weighted root mean square residual.

reported higher levels of private religious practice than Caucasians. Similarly, Hughes et al examined religious beliefs among African American and Caucasian men who were newly diagnosed with prostate cancer. They found that the older African American study participants reported higher levels of religiosity than the older Caucasian participants reported.[45]

However, none of these previous studies about the racial difference in religiosity had tried controlling for measurement bias (DIF) while making comparisons, although DIF is very common and a big concern when measures are used in health research.[46] The primary reason for the high prevalence of DIF is, while most studies in health sciences are targeting heterogeneous populations, most measures were developed from homogenous samples.[3,46] This is exactly the case of the religiosity measure adapted in this study. And under this situation, appropriate statistical methodology such as the MIMIC modeling plays a key role when comparing the population differences without producing misleading results. On the other hand, a big challenge for studies on racial differences in religiosity is the lack of culturally appropriate measurement models. As Lewis[47] noted, the lack of a culturally appropriate measure of religiosity has been a major limitation of studies examining the association between religiosity and health.

The present study filled this gap, by offering a religiosity measure adapted to diverse populations of blacks and whites with full-scale psychometric validation, and offering an appropriate methodology (MIMIC modeling) for adjustment of measurement bias (DIF) when making comparisons among racial groups.

Future studies could validate the newly adapted measure to more diverse populations other than blacks and whites, and evaluate religiosity among them with a number of health conditions that are common to older adults, such as diabetes, cardiovascular disease, and cancer. The results of these studies would broaden the applicability of this newly adapted measure, and highlight intervention targets to improve disease-related outcomes, and therefore would add significantly to the existing research literature on this topic.

4.1. Study strengths

The study has several considerable strengths. First, it offered a demonstration of measure adaptation and reporting in diverse populations, which is a premature research field.[4] Second, through careful adaptation and rigorous validation of a measure that already showed high utility, this study indicated that the adapted measure can be used for future studies examining racial differences in religiosity for diverse populations of blacks and whites, and possibly for other races if appropriate tests are passed. Given the moderate length of the adapted measure (8 items), it is neither too complicated like other religiosity measures especially in psychology, nor too simple like those used in the National Comorbidity Survey (NCS).[48] Long measures are redundant and may upset respondents in household surveys. This modified brief measure of religiosity can be well-suited for use in large surveys where the main focus is not on religion but some other health, mental health, or political outcomes. The third strength of this study is about methodology: through MIMIC models, it appropriately examined racial differences in religiosity when measurement bias (DIF) is present. The MIMIC model has an advantage in measuring group differences in latent trait when DIF is present. Although MIMIC models have been developed a few decades ago,[36] they have only recently been applied in medical research.[38,49–51] This study used a single group MIMIC model to study DIF, and used an iterative procedure. This strategy has the advantage of assessing the effect of DIF on the group differences, and avoids the shortcoming of doing too many statistical tests simultaneously.[52] While multiple group confirmatory factor analysis (MG-CFA) is another approach to study measurement invariance among different groups, we do not use it in this study, because MG-CFA requires many statistical tests simultaneously that needs a big sample size that our study cannot offer, and because MG-CFA does not offer the capacity of adjusting for DIF when comparing group differences that MIMIC has.

4.2. Study limitations

The inadequate reliability as represented by the Cronbach alphas of some of the domains may be due to the following limitations. First of all, human behavior is never static, which directly impacts the reliability of a measure in social science.[53] Second, due to simplification of the measure, the numbers of items in the domains and measure are very few (2 items in OR, 4 in NOR, 2 in SR, and 8 in overall measure). Cronbach alpha is a function of the number of items and the average correlation among the items,[54] and at a given level of average inter-item correlation, Cronbach

alpha decreases with the decrease in the number of items.[55] Third, the limited sample size (n = 196) may also contribute to the low Cronbach alphas. Some authors argue that small sample size may result in unstable estimates for Cronbach alpha,[56,57] and many authors think a sample size of 300 is "small" for Cronbach alpha estimation.[58] Fourth, Cronbach alpha is the lower bound of reliability.[55] Fifth, given that all of the 8 items are ordinal, the recently proposed ordinal version of Cronbach alpha may be calculated, which usually is larger than the traditional Cronbach alpha.[59] However, this new version of Cronbach alpha was not used in this study, mainly due to its unpopularity and calculation difficulties. On the other hand, the low reliability can also be interpreted as a trade-off for the brevity of the adapted measure. Also, sample size is not optimal in study. However, according to literatures on this issue,[35,60] it is reasonable to run CFA for whites (n = 119) and overall (n = 196), and to run MIMIC on the overall sample (n = 196).

It would be ideal to have another independent sample to cross-validate the findings of this study. Even though due to constrain of resources to this study we could not recruit such an independent sample, it is indeed an important direction for future efforts.

5. Conclusions

In summary, this study reported the adaptation and validation of a black-specific religiosity measure for diverse populations of blacks and whites, and examined racial differences of religiosity in an older population of blacks and whites. Although measure adaptation is appealing in the emerging scientific field of measurement in diverse populations, caution of measurement bias is warranted when applying the adapted measure to diverse populations. This is shown by the fact that in present study the observed racial difference in religiosity has been substantially contaminated by measurement bias (DIF), and therefore adjustment to DIF is a must when comparing racial difference of religiosity using this adapted measure. Moreover, this study illustrated the capability of the single-group MIMIC model to study population heterogeneity in a latent trait such as religiosity and to study the impact of DIF on this population difference.

Acknowledgments

References

[1] de Vet HCW, Terwee CB, Mokkink LB. Preface. Susan Ellenberg, Robert C. Elston, Brian Everitt, Frank Harrell, Jos W.R. Twisk (Series advisors). Measurement in Medicine. New York, NY: The Cambridge University Press; 2011. ix–x.

[2] Teresi JA, Stewart AL, Morales L, et al. Measurement in a multi-ethnic society: overview to the special issue. Med Care 2006;44(suppl 3):2–3.

[3] Stahl SM, Hahn AA. The National Institute on Aging's Resource Centers for Minority Aging Research. Contributions to measurement in research on ethnically and racially diverse populations. Med Care 2006;44:1–2.

[4] Stewart AL, Thrasher AD, Goldberg J, et al. A framework for understanding modifications to measures for diverse populations. J Aging Health 2012;24:992–1017.

[5] Moberg DO. Religiosity in old age. Gerontologist 1965;5:78–87.

[6] Arcury TA, Stafford JM, Bell RA, et al. The association of health and functional status with private and public religious practice among rural,

[7] Balboni TA, Balboni M, Enzinger AC, et al. Provision of spiritual support to patients with advanced cancer by religious communities and associations with medical care at the end of life. JAMA Intern Med 2013;173:1109–17.

[8] Krause N, Hayward RD. Religious involvement, practical wisdom, and self-rated health. J Aging Health 2014;26:540–58.

[9] de Camargos MG, Paiva CE, Barroso EM, et al. Understanding the differences between oncology patients and oncology health professionals concerning spirituality/religiosity: a cross-sectional study. Medicine 2015;94:e2145.

[10] Rossi M, Scappini E. Church attendance, problems of measurement, and interpreting indicators: a study of religious practice in the United States, 1975–2010. J Sci Stud Relig 2014;53:249–67.

[11] Moberg DO. Subjective measures of spiritual well-being. Rev Relig Res 1984;25:351–64 (June 1984).

[12] Moberg DO. Assessing and measuring spirituality: confronting dilemmas of universal and particular evaluative criteria. J Adult Dev 2002;9:47–60.

[13] Mindel CH, Vaughan CE. A multidimensional approach to religiosity and disengagement. J Gerontol 1978;33:103–8.

[14] Ainlay SC, Smith DR. Aging and religious participation. J Gerontol 1984;39:357–63.

[15] Chatters LM, Taylor RJ. Age differences in religious participation among black adults. J Gerontol 1989;44:S183–9.

[16] Chatters LM, Levin JS, Taylor RJ. Antecedents and domains of religious involvement among older black adults. J Gerontol 1992;47:S269–78.

[17] Taylor RJ, Chatters LM, Jayakody R, et al. Black and white differences in religious participation: a multisample comparison. J Sci Stud Relig 1996;35:403–10.

[18] Levin JS, Chatters LM. Religion, health, and psychological well-being in older adults: findings from three national surveys. J Aging Health 1998;10:504–31.

[19] Taylor RJ, Chatters LM, Joe S. Non-organizational religious participation, subjective religiosity, and spirituality among older African Americans and Black Caribbeans. J Relig Health 2011;50:623–45.

[20] Krause N. A comprehensive strategy for developing closed-ended survey items for use in studies of older adults. J Gerontol B Psychol Sci Soc Sci 2002;57:S263–74.

[21] Fetzer I. Multidimensional Measurement of Religiousness/Spirituality for Use in Health Research: A Report of the Fetzer Institute/National Institute on Aging Working Group. Kalamazoo, MI:John E. Fetzer Institute; 2003.

[22] Chatters LM, Taylor RJ, Lincoln KD. Advances in the measurement of religiosity among older African Americans: implications for health and mental health researchers. J Mental Health Aging 2001;7:181–200.

[23] Levin JS, Markides KS. Religious attendance and subjective health. J Sci Stud Relig 1986;25:31–40.

[24] Levin JS, Vanderpool HY. Is frequent religious attendance really conductive to better health? Soc Sci Med 1987;24:589–600.

[25] Levin JS, Chatters LM, Taylor RJ. Religious effects on health status and life satisfaction among black Americans. J Gerontol 1995;50B:S154–63.

[26] Bot SDM, Terwee CB, van der Windt DAWM, et al. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis 2004;63:335–41.

[27] DeVellis RF. Redundancy. Acquisitions Editor: Vicki Knight; Editorial Assistant: Kalie Koscielak; Production Editor: Astrid Virding; Copy Editor: Megan Speer; Permissions Editor: Adele Hutchinson. Scale Development: Theory and Applications 3rd ed. Los Angeles, CA: The SAGE Publications, Inc.; 2012. 77–80.

[28] Reeve BB. Special issues for building computerized-adaptive tests for measuring patient-reported outcomes: the National Institute of Health's investment in new technology. Med Care 2006;44(suppl 3):S198–204.

[29] Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. Value Health 2009;12:419–29.

[30] Selman L, Harding R, Gysels M, et al. The measurement of spirituality in palliative care and the content of tools validated cross-culturally: a systematic review. J Pain Symptom Manage 2011;41:728–53.

[31] DeVellis RF. Evaluate the Items. Scale Development: Theory and Applications 3rd ed.Los Angeles, CA:The SAGE Publications, Inc.; 2012. 104–10.

[32] Yang FM, Heslin KC, Mehta KM, et al. A comparison of item response theory-based methods for examining differential item functioning in object naming test by language of assessment among older Latinos. Psychol Test Assess Model 2011;53:440–60.

ethnically diverse, older adults with diabetes. J Rural Health 2007; 23:246–53.

[33] Brown TA. Confirmatory Factor Analysis for Applied Research. New York, NY:The Guilford Press; 2006.

[34] Muthén LK, Muthén BO. Mplus Short Courses: Traditional Latent Variable Modeling with Mplus. Los Angeles, CA: Muthén & Muthén; 2003. Available at: http://www.statmodel.com/course_materials.shtml. Accessed August 1, 2016.

[35] Kline RB. Principles and Practice of Structural Equation Modeling. 2nd ed.New York, NY:The Guilford Press; 2005.

[36] Bollen KA. Structural Equations with Latent Variables. New York, NY: John Wiley & Sons; 1989.

[37] Muthén LK, Muthén BO. Mplus User's Guide. 7th ed.Los Angeles, CA: Muthén & Muthén; 1998–2015.

[38] Jones RN. Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination: detecting differential item functioning using MIMIC modeling. Med Care 2006;44(suppl 3):S124–33.

[39] Zumbo BD. Rao CR, Sinharay S. Validity: Foundational Issues and Statistical Methodology. Handbook of Statistics [26], Psychometrics. Amsterdam, The Netherlands:Elsevier; 2007. 45–79.

[40] Mellenbergh GJ. Item bias and item response theory. Int J Educ Res 1989;13:127–43.

[41] Teresi JA, Kleinman M, Ocepek-Welikson K. Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. Stat Med 2000;19:1651–83.

[42] Krause N. Exploring race and sex differences in church involvement during late life. Int J Psychol Relig 2006;16:127–44.

[43] Taylor RJ, Chatters LM, Jackson JS. Religious and spiritual involvement among older African Americans, Black Caribbeans and Whites: findings from the National Survey of American Life. J Gerontol 2007;62B: S238–50.

[44] Taylor RJ, Chatters LM. Religious media use among African Americans, Black Caribbeans, and Non-Hispanic Whites. J Afr Am Stud 2011;15:433–54.

[45] Hughes HC, Barg FK, Weathers B, et al. Differences in cultural beliefs and values among African American and European American men with prostate cancer. Cancer Control 2007;14:277–84.

[46] Teresi JA, Fleishman JA. Differential item functioning and health assessment. Qual Life Res 2007;16:33–42.

[47] Lewis LM. Spiritual Assessment in African-Americans: a review of measures of spirituality used in health research. J Relig Health 2008; 47:458–75.

[48] Kessler RC, Galea S, Jones RT, et al. Mental illness and suicidality after Hurricane Katrina. Bull World Health Organ 2006;84:930–9.

[49] Jones RN, Gallo JJ. Education and sex differences in the mini-mental state examination: effects of differential item functioning. J Gerontol 2002;57B:548–58.

[50] Jones RN. Racial bias in the assessment of cognitive functioning of older adults. Aging Mental Health 2003;7:83–102.

[51] Yu YF, Yu AP, Ahn J. Investigating differential item functioning by chronic disease in the SF-36 health survey: a latent trait analysis using MIMIC models. Med Care 2007;45:851–9.

[52] Teresi JA. Different approaches to differential item functioning in health applications: advantages, disadvantages and some neglected topics. Med Care 2006;44(suppl 3):S152–70.

[53] Merriam SB. Merriam SB. Reliability or Consistency. Qualitative Research: A Guide to Design and Implementation. San Francisco, CA: Josse-Bass; 2009. 220–3.

[54] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.

[55] Carmines EG, Zeller RA. Reliability and Validity Assessment. Thousand Oaks, CA:SAGE Publications Ltd; 1979.

[56] Charter RA. Study samples are too small to produce sufficiently precise reliability coefficients. J Gen Psychol 2003;130:117–29.

[57] Yurdugül H. Minimum sample size for Cronbach's coefficient alpha: a Monte-Carlo study. Hacettepe Univ J Educ 2008;35:397–405.

[58] Segall DO. The reliability of linearly equated tests. Psychometrika 1994;59:361–75.

[59] Zumbo BD, Gadermann AM, Zeisser C. Ordinal versions of coefficients alpha and theta for Likert rating scales. J Mod Appl Stat Methods 2007;6:21–9.

[60] Iacobucci D. Structural equations modeling: fit indices, sample size, and advanced topics. J Consum Psychol 2010;20:90–8.