Differential Item Functioning on the Schedule of Racist Events: Exploring a Modified Scale for Samples with Black and White Participants

Richard Feinn, PhD; Kristina P. Schumann, PhD; Howard Tennen, PhD; Julie Wagner, PhD

Objective: Although both Black and White individuals report racial discrimination, self-report measures of exposure to racial discrimination that can be used across races/ethnicities are rare. The primary aim of our study was to determine if the Schedule of Racist Events (SRE), which was designed for use in Black samples, should also be used in White samples, and if so, what modifications to the scale are necessary.

Methods: In a sample of 302 adults, approximately equally divided by race, we investigated whether item endorsement differed between Black and White respondents.

Results: Results of confirmatory factor analysis and differential item functioning (DIF) analysis suggest that changing the item stem (from 'because you are Black' to 'because of your race/ethnicity') and removing four items that show differential item functioning and/or do not load on the first factor, results in a psychometrically sound scale with no evidence of measurement bias.

Conclusion: Researchers interested in measuring racial discrimination in samples that include both Black and White respondents may consider using this version of the SRE. Future studies should investigate other forms of validity in Black and White samples. (*Ethn Dis*. 2014;24[4]:406–412)

Key Words: DIF, SRE, Perceived Racism

From Division of Behavioral Sciences and Community Health, University of Connecticut Health Center (JW); and Frank H. Nedder, M.D. School of Medicine, Quinnipiac University (RF); and VA Connecticut Healthcare System, West Haven, Connecticut (KS); and Department of Community Medicine & Healthcare, University of Connecticut Health Center (HT).

Address correspondence to Julie Wagner, PhD; Division of Behavioral Sciences and Community Health; MC3910 University of Connecticut Health Center; 263 Farmington Ave; Farmington, CT 06030; 860.679.4508; 860.679.1342(fax); juwagner@uchc.edu

Introduction

Growing evidence points to a deleterious relationship between self-reported racial discrimination and greater psychological symptoms, poorer physical health[1] and increased mortality risk.[2] Although most research has focused on the effects of self-reported racial discrimination in Blacks, there is evidence that these relations exist in other racial/ethnic groups, including Whites. Most studies that include Whites show that they, too, report racial discrimination, albeit at much lower rates than Blacks.[3,4] There is also evidence that, for some outcomes, the relationship between discrimination and outcomes may vary by race, and not always in expected directions. For example, Barnes and colleagues[2] reported that self-reported racial discrimination was related to increased risk of mortality, and further reported that this relation was, contrary to their expectations, stronger in Whites than Blacks. Thus, self-reported racial discrimination is a novel risk factor for health and mental health outcomes that is open to investigation in both Blacks and Whites.

A variety of scales have been designed to measure exposure to racial discrimination (see Williams & Mohammed for a comprehensive review).[5] To our knowledge, only one scale has been designed to measure racial discrimination across race/ethnicity – the Experiences of Discrimination scale (EOD).[6] Ideally, when measuring any construct, investigators should use a multi-method approach. When only one scale is available, any idiosyncrasies of the scale may be ascribed to the construct, leading to potentially erroneous conclusions about the construct. Thus, it is important that additional measures of racial discrimination be validated for use across race/ethnicity.

The Schedule of Racist Events (SRE),[7] which was designed for use in Black samples, assesses the respondent's exposure to and appraisal of instances of racial discrimination in a variety of domains in which the respondent is the direct and individual target. The SRE has led to important insights regarding the nature and effects of racial discrimination. For example, investigators using the SRE have found that self-reported racial discrimination was positively associated with health risk behaviors[8] and poorer self-rated health.[9] Further, Klonoff and Landrine demonstrated, using the SRE, that light-skinned Blacks report less exposure to racial discrimination than dark-skinned Blacks.[10]

Whereas the SRE was designed for, and has proven useful in, Black samples, its use beyond Blacks has not been investigated. We were interested in examining whether it should be used with Whites. Because much of the research on the effects of racist events pertains to health outcomes, we examined this question in a clinical sample of individuals with diabetes – a disease determined, in part, by psychosocial influences and with well-established racial and ethnic health disparities. The primary aim of our study was to establish if the SRE demonstrates measurement bias in item endorsement contributing to differences between Blacks' and Whites' responses beyond racial discrimination *per se*. To examine this question, we conducted differential item functioning (DIF) analysis. Ideally, responses to items within an an instrument should only be influenced by the construct the instrument is designed to measure, so that any difference in item

The primary aim of our study was to establish if the SRE demonstrates measurement bias in item endorsement contributing to differences between Blacks' and Whites' responses beyond racial discrimination per se.

responses between groups is solely due to differences in the level of the construct. Differential item functioning is present when other factors besides the construct of interest contribute to how a person responds to an item. This can lead to two individuals from different groups who are the same in the level of the construct to have a different probability of endorsing an item. When DIF is present, the instrument cannot be used to make comparisons between groups because the scale score that results does not necessarily reflect the level of the attribute similarly across groups. We examined evidence of its validity in samples of Blacks and Whites.

METHODS

Sample and Recruitment

Our study involved secondary data analysis from several of our studies on psychosocial aspects of diabetes. The sample of convenience reported here was combined from three smaller studies. Whereas recruitment strategies varied by study, each administered identical measures, and applied identical criteria for language, age, and race/ethnicity. Inclusion criteria included English speaking, aged ≥18 years, and self-reported diabetes. For Blacks, inclusion criteria included having two parents of African descent, being born and raised in the United States, and self-identifying as Black or African American. For Whites, inclusion criteria included having two parents of European descent, being born and raised in the United States, and self-identifying as White or Caucasian. In each of the three studies, trained research assistants were available to administer the questionnaires when needed and answer questions. All procedures were approved by the University of Connecticut Health Center institutional review board.

One-hundred and two participants were part of a community based study investigating racial discrimination in persons with diabetes. Participants included a convenience sample of adults with diabetes attending the health fairs sponsored by the American Diabetes Association in the northeast United States from 2004–2006. The investigator (JW) hosted a booth where Black attendees with diabetes were invited to participate in the study.

A smaller number (n=20) participated in a clinic based study investigating racial discrimination and oral health at the University of Connecticut Health Center general dentistry clinic in 2009. Black participants were recruited through flyers posted in the waiting room of the general dentistry clinic.

The remainder (n=180) participated in a study comparing psychosocial characteristics of women with diabetes to women without diabetes between 2004 and 2009 at the University of Connecticut Clinical Research Center. Participants were recruited through Connecticut newspapers, electronic mail, paycheck inserts and local medical practices. Recruits were excluded if they had acute medical or psychiatric problems requiring intensive outpatient treatment or hospitalization, current drug or alcohol use disorder.

The first two studies were exclusively Black samples and the last study included more Whites (70%) than Blacks (30%). There was no difference in SRE scores among Blacks from the three studies.

Measures

Racial Discrimination

Self-reported lifetime exposure to racial discrimination was assessed with the SRE.[7] Eighteen items measure the frequency of overt and subtle racist interpersonal situations in which the respondent is the direct and individual target of prejudice and/or discrimination. Items include ''How many times have you been treated unfairly by your co-workers, fellow students, or colleagues because you are Black?'' Because we were investigating racist events among both White and Black women, we modified the items to read, ''…because of your race or ethnicity.'' We also modified the item ''How many times have you been called a racist name like…or other names?'' to read, ''How many times have you been called a racist name?''

Respondents are asked to complete the same items three times, under three different sets of instructions - lifetime exposure, recent exposure (past year), and appraisal of the exposure (subjective stressfulness). Thus, three responses are produced for each item. The original study that examined the factor structure of the SRE reported three distinct subscales that each formed a single factor 'racist events.'[11] The first subscale measures lifetime exposure to racial discrimination. For example, ''How often in your entire life have you been treated unfairly by teachers because you are Black?'' The second subscale measures recent exposure. For example, ''How often in the past year have you been treated unfairly by teachers because you are Black?'' For both, respondents reply on a 6-point scale ranging from 1 (the event never happened to me) to 6 (the event happened almost all of the time). The third subscale measures the subjective stressfulness of each endorsed exposure with the question, ''How stressful was this for you?'' on a 6-point scale ranging from 1 (not at all stressful) to 6 (extremely stressful). An 18th item

asks "How different would your life be had you not been treated in a racist and unfair way?" Respondents reply on a 6-point scale ranging from 1 (same as now) to 6 (totally different).

Findings for the three subscales, each measuring a single construct, were generally replicated in a sample of $n=246$ Black women[12] and $n=323$ Black undergraduate students.[11,13] These studies noted very high inter-correlations among subscales.[13,14] The high inter-correlations are likely because responses to items in the third subscale (stressfulness) are dependent upon responses to items in the other subscales, which are themselves inextricably linked. This substantial overlap has led investigators to compute SRE scores in a variety of ways, combining (summing or multiplying) different permutations of the subscales or relying on selected items for study-specific needs. Investigators have recommended that the scale be modified to reduce redundancy.[13,14] It is the psychometric properties of the first subscale – lifetime frequency – that forms the basis of the rest of the scale, and the psychometric properties of which are of critical importance and are therefore the focus of this report.

Demographic and Clinical Characteristics

In fully adjusted analyses, we controlled for demographics. Demographic information included age, sex, and marital status. Socioeconomic status (SES) indicators included education and income, and clinical characteristics included smoking status and body mass index (BMI).

Data Analysis

Descriptive statistics were used to summarize demographic variables by race and included sex, education and income level, and marital status. Independent samples *t*-tests were used to compare continuous measures and chi-square test for categorical variables.

Prior to testing for DIF, exploratory factor analysis was performed to investigate the dimensionality of the SRE frequency scale. We first wanted to establish that a single underlying construct could explain the covariation among items since this is a necessary prerequisite for a DIF analysis. This was established by examining the difference among successive eigenvalues and the factor loadings from single and multiple factor models. We then followed this with a confirmatory factor analysis, which was driven by the exploratory results, and compared competing models by fit indices to arrive at the best unidimensional model. Model fit information was based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), the incremental measure comparative fit index (CFI), and the absolute measure root mean square error of approximation (RMSEA).

After finding the best unidimensional model from the factor analyses, DIF analysis was conducted using a multiple indicator multiple causes (MIMIC) structural equation model. The MIMIC model is able to reveal uniform differential item functioning. Uniform DIF exists when conditioning or equating on the latent construct, the probability of a certain response differs by group membership throughout the entire range of the construct. In other words, a Black and a White individual who are estimated to have experienced the same level of racial discrimination still respond differently to the item. The MIMIC model has been shown to perform well in situations in which the sample size is not very large,[15] as in the present study.

To test for DIF, a measurement model was created in which all items were indicators of the latent construct, racial discrimination, and the construct was regressed on an indicator variable, race. The loadings of each item on the construct racial discrimination represent the measurement model. The construct is regressed on race and this relationship is captured by the loading. Each item in turn was regressed on race (binary indicator 1=Black, 0=White) and the path from race to the item measured the direct effect of race beyond the effect of race through the latent construct racial discrimination. The test of this path is a test for DIF and represents the difference in response beyond what can be attributed to racial discrimination. Two models were run, an unadjusted model as previously described and an adjusted model. The adjusted model included as covariates any demographic variables that significantly differed by race. To control the type I error rate due to testing numerous items, the Benjamini and Hochberg (BH) false discovery rate[16] procedure was applied.

Whereas the range of item responses was six, the frequency distribution for most items was leptokurtic and positively skewed. Often 50% or more of the respondents selected the option never and for some items almost always was not chosen. Thus for all analyses items were treated as ordinal and a robust maximum likelihood estimator was used.[17] Factor analysis and DIF were carried out using Mplus version 7.[18]

To complement the MIMIC model an ordinal logistic regression model was also run, which is able to test for non-uniform DIF.[19] In the logistic regression the total SRE score, race, and the interaction between total score and race are entered in a model for each item. A significant interaction between total score and race is evidence for non-uniform DIF and means differences between the races on endorsing the item depend on the total level of racist events. As with the MIMIC model, significant demographic variables were included as covariates and the false discovery rate was applied.

RESULTS

Sample

The final sample, which was composed of three smaller studies, consisted of 302 participants of which 58% self-

Table 1. Participant demographic characteristics by race[a]

Characteristic	Black (*n*=176)	White (*n*=126)	*P*
Age, mean ± SD	54.7 ± 10.9	60.5 ± 9.0	<.001
Sex, female	140 (79.5)	126 (100)	<.001
Education			.002
≤High school	51 (29.3)	20 (15.9)	
Some college	68 (39.1)	44 (34.9)	
4 Year college	30 (17.2)	24 (19.0)	
Graduate	25 (14.4)	38 (30.2)	
Income			<.001
$0–20,000	54 (33.5)	17 (13.5)	
$21–40,000	43 (26.7)	24 (19.0)	
$41–60,000	33 (20.5)	21 (16.7)	
$61–80,000	14 (8.7)	18 (14.3)	
>$80,000	17 (10.6)	46 (36.5)	
Marital status			<.001
Married/cohabitating	55 (31.2)	72 (57.1)	
Other	121 (68.8)	54 (42.9)	
Current smoker, yes	14 (8.0)	3 (2.4)	.035
BMI, mean ± SD	33.7 ± 8.2	31.1 ± 7.1	.004

[a] Data are n (%) unless indicated otherwise.

identified as Black and 42% as White. As displayed in Table 1, all demographic variables differed by race. Compared to Whites, Blacks were younger, included more males, had a lower educational attainment and income level, were less likely to be married, smoked at a higher rate, and were heavier as measured by BMI. Thus, DIF analyses controlled for age, sex, marital status, education and income.

Table 2 shows the mean and standard deviation for each item by race. In general the item means were close to 1 or 2 (once in a while) and scores indicate prevalent, but infrequent, racial discrimination. For every item, except two, Blacks reported significantly more racial discrimination than Whites ($P<.01$); the two exceptions were being angry (item 13) and getting into an argument (item 16) ($P>.05$). Angry was the highest item mean for Whites (2.46), whereas argument was the only item of the 18 for which Whites reported a higher mean than Blacks (1.73 vs 1.59 respectively). Also, there was a large difference of over one standard deviation in total score between Black and White respondents.

Table 2. SRE item descriptive statistics[a]

	Black	White
SRE Item	Mean (SD)	Mean (SD)
1 Teachers	1.95 (1.11)	1.09 (.33)
2 Employers	2.36 (1.28)	1.19 (.51)
3 Coworkers	2.02 (1.07)	1.27 (.68)
4 Service	2.35 (1.12)	1.17 (.52)
5 Strangers	2.26 (1.06)	1.28 (.58)
6 Helping	1.74 (.92)	1.17 (.57)
7 Neighbors	1.69 (.97)	1.14 (.60)
8 Institutions	1.91 (1.02)	1.04 (.19)
9 Friends	1.58 (.83)	1.12 (.32)
10 Accused	1.79 (.96)	1.10 (.62)
11 Misunderstood	2.16 (1.12)	1.18 (.50)
12 Tell off	2.72 (1.34)	2.23 (1.17)
13 Angry	2.80 (1.29)	2.46 (1.18)
14 Drastic	1.33 (.70)	1.05 (.22)
15 Names	2.00 (1.10)	1.27 (.64)
16 Argument	1.59 (.87)	1.73 (.91)
17 Made fun of	1.51 (.82)	1.21 (.49)
18 How different	2.51 (1.56)	1.35 (.83)
Total	36.2 (12.0)	24.0 (6.3)

SRE, schedule of racist events.
[a] Item range is 1 to 6, with 6 indicating more experience of discrimination.

Factor Analysis

An exploratory factor analysis resulted in a value of 10.0 for the first eigenvalue and 1.3 for the second eigenvalue, with the remaining eigenvalues below 1.0. We thus focused on the one and two factor models for comparisons. The one factor model resulted in all 18 items loading, with an AIC value of 9884 and a BIC value of 10262. The two factor model redistributed items 12, 13, and 16 onto a second factor and they did not significantly load on the first factor. In the single factor model these three items had the lowest loadings (see Table 3). Using Geomin rotation[20], the two factors were reasonably correlated, $r=.53$. The fit indices were lower than for the one factor model, indicating a better fit, with an AIC value of 9781 and a BIC value of 10223. We reran the one factor model excluding items 12, 13, and 16 one at a time and together; excluding all three resulted in the lowest AIC (7492) and BIC (7804).

We next performed confirmatory factor analyses and used the RMSEA and CFI to assess model fit. The one factor model using all 18 items resulted in a good CFI of .976 and an acceptable RMSEA of .074, whereas the fit of the two factor model was slightly better; CFI=.99 and RSMEA=.046. A one factor model excluding items 12, 13, and 16 also fit better than the 18 item one factor model; CFI=.992 and

Table 3. One and two factor model factor loadings for exploratory factor analysis with Geomin rotation of the Schedule of Racist Events frequency scale

SRE Item	One Factor Model	Two Factor Model	
		1	2
1 Teachers	.79[a]	.92[a]	−.20
2 Employers	.79[a]	.79[a]	.01
3 Coworkers	.81[a]	.81[a]	.01
4 Service	.83[a]	.84[a]	.00
5 Strangers	.80[a]	.82[a]	−.02
6 Helping	.81[a]	.78[a]	.07
7 Neighbors	.78[a]	.78[a]	−.01
8 Institutions	.83[a]	.81[a]	.07
9 Friends	.75[a]	.84[a]	−.12
10 Accused	.83[a]	.80[a]	.05
11 Misunderstood	.83[a]	.82[a]	.04
12 Tell Off	.54[a]	.21	.59[a]
13 Angry	.48[a]	.01	.85[a]
14 Drastic	.58[a]	.64[a]	−.10
15 Names	.69[a]	.63[a]	.13
16 Argument	.34[a]	−.01	.61[a]
17 Made fun of	.61[a]	.67[a]	−.08
18 How different	.74[a]	.68[a]	.10

[a] $P<.05$.

RSMEA=.049. A corrected chi-square difference test[21] between the single factor model with all 18 items and the single factor model with 15 items factor resulted in a superior fit for the 15 item model ($\chi^2_{(45)} = 202$, $P<.001$). From the above results it was decided the DIF analysis should exclude items 12, 13, and 16 because the three items did not appear to measure the same underlying construct as the remaining 15 items.

Differential Item Functioning

The results from the MIMIC model DIF are presented in Table 4. The first column lists the item and the next two columns are the unadjusted results. The standardized coefficient is the effect of race on that particular item (β) and the next column is the BH corrected *P*. One of the 15 items showed evidence of statistically significant DIF. Item 17, made fun of, had a negative coefficient, indicating that Blacks required a higher threshold for endorsement. The next two columns of table 4 show the results after adjusting for the demographic variables age, sex, educational level, income level, marital status, smoking status, and BMI. After adjustments for demographic differences, item 17 remained statistically significant ($P=.015$) while the other items remained non-significant.

The ordinal logistic regression, testing for non-uniform DIF, resulted in no statistically significant interactions (all $P>.05$) between SRE total score and race after controlling for demographic variables, However, the lowest *P* (.06) was for item 17, made fun of, which was also identified as demonstrating DIF through the MIMIC model. The negative coefficient (b = −.18) for the interaction indicates that as the level of total perceived discrimination increases Blacks are less likely to endorse the item compared to Whites.

Table 4. Multiple Indicator Multiple Cause Model differential item functioning results

SRE Item	Unadjusted		Adjusted for Demographics	
	Standardized B	*P*	Standardized B	*P*
1 Teachers	.08	.467	.11	.302
2 Employers	.12	.095	.15	.080
3 Coworkers	−.12	.173	−.08	.401
4 Service	.11	.113	.06	.477
5 Strangers	.00	.934	.07	.420
6 Helping	−.02	.778	−.05	.724
7 Neighbors	−.11	.396	−.05	.786
8 Institutions	.17	.095	.21	.080
9 Friends	−.08	.484	−.13	.302
10 Accused	.09	.467	−.01	.943
11 Misunderstood	−.10	.199	−.19	.114
14 Drastic	.10	.605	−.02	.943
15 Names	−.14	.173	−.19	.114
17 Made fun of	−.32	.008	−.35	.015
18 How different	−.04	.661	.01	.943

DISCUSSION

The main finding from our study is that, with some minor modifications, the SRE can be used successfully in White as well as Black samples. Results of confirmatory factor analysis and DIF analysis suggest that modifying some items and removing others results in a psychometrically sound scale. Specifically, when used with Whites, we recommend changing the item stem (from 'because you are Black' to

The main finding from our study is that, with some minor modifications, the SRE can be used successfully in White as well as Black samples.

'because of your race/ethnicity'), modifying item 15 (from "How many times have you been called a racist name like…or other names?" to "How many times have you been called a racist name?") and removing four items (12, 13, 16, and 17). Researchers interested in measuring racial discrimination in Blacks and Whites using the same scale may consider using this race and ethnicity-neutral version of the SRE.

In the original examination of the factor structure of the SRE, Landrine and Klonoff[7] reported that, for each of the three subscales (recent frequency of racist events, lifetime frequency of racist events, and appraisal of racist events), a single factor, containing all items, should be retained. Findings from the current study indicate that, in this sample of Black and White persons with diabetes, the lifetime frequency of racist events subscale is best described as having two factors, or one abbreviated factor (15 items rather than 18). Items that did not load well on the single factor included "How many times [in your entire life] did you want to tell someone off for being racist, but didn't say anything?", "How many times [in your entire life] have you been really angry about something racist?" and "How many times [in your entire life] have you gotten into an argument or a fight about something racist that was done to you or done to someone else?" A careful review of the wording of these items reveals that they reflect reactions to a racist event rather than exposure to a racist event per se. Unlike the other items in the scale, these items also do not explicitly make the respondent the direct and individual target of the racial discrimination. We suggest that these items may measure a sensitivity to anger in response to others' unfair treatment, rather than or in addition to, one's own unfair treatment. In this way, Whites, who by their own reports are less frequently the direct and intended target of racial discrimination than are Blacks, may still have scores on these items equal to or above those of Blacks. This interpretation is supported by the means on these items. The "tell off" and "angry" items had similar means for Blacks and Whites, and "argument" actually had a higher mean score for Whites than Blacks. Therefore, we recommend that researchers specifically interested in assessing exposure to racist events in which the respondent is the individual and direct target of prejudice and/or discrimination should omit these 3 items, or consider them as a separate subscale.

The DIF results from our study indicate that the SRE lifetime frequency of racist events subscale generally functioned similarly for Blacks and Whites with the exception of one item. Whites had a lower threshold for endorsing item 17 – made fun of, picked on, pushed, shoved, hit, or threatened with harm [because of your race/ethnicity]. Consistent with standard DIF interpretation, we recommend eliminating items that demonstrate significant DIF when using the scale in samples that include both Black and White participants.[22,23] We are not recommending removing items based on significant DIF when the scale is used with exclusively Black samples.

When working with a measure such as the SRE, it is important to note the distinction between racist events and racism. Racist events are understood here to be discrete acts of racially motivated prejudice and discrimination. Racism, on the other hand, is the broader societal context of unequal access to power and privilege. Thus, while White participants report experiences of prejudice and discrimination, we do not suggest here that they are victims of racism per se. It is for this reason – the existence of systematic and institutional racism imbedded in society—that the higher rates of exposure could be expected from Black respondents.

Our study provides evidence for the SRE's construct validity and lack of measurement bias between Black and White samples. Other properties should also be examined. Future research should examine other forms of validity including convergent, divergent, and criterion validity. If those properties are demonstrated, then investigators can feel confident using the SRE with White, as well as Black, samples.

Limitations

This study has limitations. First, our sample size of 302 participants is considered small for a differential item function analysis with approximately 20 observed variables. A simulation study[24] to detect uniform DIF using ordinal items, such as the SRE, suggested 200 participants per group, whereas we had approximately 150 per group. Researchers who have used the SRE with greater number of participants should attempt to replicate our results to help assess the validity of our findings.

Second, our study sample was derived by combining several smaller studies that differed on demographic variables. The resulting sample was predominately (90%) women, so sex differences could not be assessed and it is unclear if the current findings are generalizable to men. Klonoff and Landrine[11] demonstrated sex differences on the SRE, with men reporting greater frequency (both recent and lifetime) of racist events. However, the authors noted that these differences were small and may not be meaningful in understanding whether men and women experience racism differently. Additionally, in the cross-validation study, item-by-item analyses of sex differences were not conducted, so it is not known whether specific items behave differently for men and women. Future research should include a mixed-sex sample. These limitations are generally outweighed by the study's strengths including the extension of racial discrimination to a sample not exclusively Black, the careful psychometric evaluation of a common measure, and control of demographic factors. If findings are replicated, the modified SRE described here may be a

useful measure for understanding the experience of racial discrimination in samples comprised of both White and Black participants.

ACKNOWLEDGMENTS

This research was supported by grants to Dr. Wagner from the National Institutes of Diabetes, Digestive, and Kidney Diseases (R21 DK074468), the American Diabetes Association (7-05-CR-31), the American Heart Association (0535301N), and the Uconn Health Center General Clinical Research Center (M01RR006192).

REFERENCES

1. Pascoe AE, Richman LS. Perceived discrimination and health: A meta-analytic review. *Psychol Bull*. 2009;135:531–554.
2. Barnes LL, de Leon CF, Lewis TT, Bienias JL, Wilson RS, Evans DA. Perceived discrimination and mortality in a population-based study of older adults. *Am J Public Health*. 2009;98(7):1241–1247.
3. Borrell LN, Kiefe CI, Diez-Roux AV, Williams DR, Gordon-Larsen P. Racial discrimination, racial/ethnic segregation, and health behaviors in the CARDIA study. *Ethnic Health*. 2013;18(3):227–243.
4. Hausmann LR, Jeong K, Bost JE, Ibrahim SA. Perceived discrimination in health care and health status in a racially diverse sample. *Med Care*. 2008;46:905–914.
5. Willams DR, Mohammed SA. Discrimination and racial disparities in health: Evidence and needed research. *J Behav Med*. 2009;32:20–47.
6. Krieger N, Smith K, Naishadham D, Hartman C, Barbeau EM. Experiences of discrimination: Validity and reliability of a self-report measure for population health research on racism and health. *Soc Sci Med*. 2005; 61:1576–1596.
7. Landrine H, Klonoff EA. The Schedule of Racist Events: A measure of racial discrimination and a study of its negative physical and mental health consequences. *J Black Psychol*. 1996;22:144–168.
8. Kwate NA, Valdimarsdottir HB, Guevarra JS, Bovbjerg DH. Experiences of racist events are associated with negative health consequences for Black women. *J Natl Med Assoc*. 2003;95(6):450–460.
9. Williams DR, Neighbors HW, Jackson JS. Racial/ethnic discrimination and health: Findings from community studies. *Am J Public Health*. 2003;92(2):200–208.
10. Klonoff EA, Landrine H. Is skin color a marker for racial discrimination? Explaining the skin color-hypertension relationship. *J Behav Med*. 2000;23(4):329–338.
11. Klonoff EA, Landrine H. Cross-validation of the Schedule of Racist Events. *J Black Psychol*. 1999;25:231–254.
12. DaBlaere C, Moradi B. Structures of the Schedules of Racist and Sexist Events: Confirmatory factor analyses of Black women's responses. *Psychol Women Quart*. 2008;32:83–94.
13. Greer TM. A structural validation of the Schedule of Racist Events. *Meas Eval Couns Dev*. 2010;43(2):91–107.
14. Fischer AR, Shaw CM. African Americans' mental health and perceptions of racist discrimination: The moderating effects of racial socialization experiences and self-esteem. *J Coun Psychol*. 1999;46(3):395–407.
15. Woods C. Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivar Behav Res*. 2009;44:1–27.
16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289–300.
17. Brown T. *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press; 2006.
18. Muthen LK, Muthen BO. *Mplus User's Guide.* 7th ed. Los Angeles, CA: Muthen & Muthen; 2012.
19. Shimizu Y, Zumbo BD. A logistic regression for differential item functioning primer. *JLTA*. 2005;7:110–124.
20. Browne WJ, Goldstein H, Rasbash J. Multiple membership multiple classification (MMMC) models. *Stat Model*. 2001;1(2): 103–124.
21. Byrne B. *Structural Equation Modeling with Mplus*. New York: Taylor & Francis Group; 2012.
22. Angoff WH. Perspectives on differential item functioning methodology. In: Holland PW, Wainer H, eds. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Earlbaum Associates Inc.; 1993;3–23.
23. Burton E, Burton NW. The effect of item screening on test scores and test characteristics. In: Holland PW, Wainer H, eds. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.; 1993; 321–335.
24. Scott N, Frayers P, Blackronson N, et al. A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *J Clin Epidemiol*. 2009;62:288–295.

AUTHOR CONTRIBUTIONS

Design and concept of study: Feinn, Schumann, Tennen, Wagner
Acquisition of data: Wagner, Tennen
Data analysis and interpretation: Feinn, Schumann, Tennen, Wagner
Manuscript draft: Feinn, Schumann, Tennen, Wagner
Statistical expertise: Feinn
Supervision: Feinn, Wagner