The Symbolic Racism Scale and the Modern Sexism Scale: Testing Measurement Invariance in Diverse Samples

Tangier M. Davis[1], Leanna J. Papp[1, 2], Majel R. Baker[3], and Sara I. McClelland[1, 2]
[1] Department of Psychology, University of Michigan
[2] Department of Women's and Gender Studies, University of Michigan
[3] Department of Psychology, College of Saint Benedict and Saint John's University

Objectives: The present study examined the measurement invariance of the Symbolic Racism Scale (SRS) and the Modern Sexism Scale (MSS) across racial/ethnic and gender groups. Previous psychometric evaluations of the SRS and MSS scores have not examined the equivalence across racial/ethnic and gender groups or have been otherwise statistically inadequate. Therefore, this study sought to fill this gap. *Method:* To establish measurement equivalence across racial/ethnic (Black, Latinx, and white) and gender (women and men) groups, we conducted a measurement invariance analysis of the SRS and the MSS in a large, diverse sample ($N = 719$). *Results:* We found that the SRS and MSS were invariant across gender, and the SRS was invariant across racial/ethnic groups. However, the MSS was noninvariant across racial/ethnic groups. Partial invariance testing revealed nonequivalent factor loadings between Black and Latinx participants compared to white participants on an item of the MSS that referenced "unwarranted" attention that women receive from the government and media. *Conclusions:* Researchers should consider reevaluating the item that reads: "Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences." Future research is needed to assess how the item is interpreted by Black and Latinx people so it can be modified for use in these communities. Our findings underscore the importance of assessing the validity of the scores in commonly used scales across diverse groups.

Public Significance Statement

The Symbolic Racism Scale (SRS) and the Modern Sexism Scale (MSS) are measures used to assess covert prejudice toward Black people and women, respectively. These scales have been used to understand groups' attitudes toward social movements, social issues, and political candidates and have previously been found to predict attitudes toward race- and gender-related policies. However, to compare group scores on psychological measures, researchers must know if the measured construct has the same structure or meaning across groups. We examined this question in the SRS and MSS by assessing the invariance of these measures across diverse gender and racial/ethnic groups. Measurement invariance indicates that the same construct was measured in each group, while noninvariance suggests that there may be measurement error for at least one of the groups we examined. We found that the MSS is noninvariant between racial/ethnic groups; specifically, an item regarding special favors for women may not be a reliable indicator of modern sexist beliefs among Black and Latinx individuals. Our findings highlight the need for more work on how modern sexism manifests within diverse communities.

Keywords: measurement invariance, Black, Latinx, symbolic racism, modern sexism

Supplemental materials: https://doi.org/10.1037/cdp0000618.supp

Tangier M. Davis ⬡ https://orcid.org/0000-0002-6870-4835
Leanna J. Papp ⬡ https://orcid.org/0000-0002-3333-9903
Majel R. Baker ⬡ https://orcid.org/0000-0001-5023-900X
Sara I. McClelland ⬡ https://orcid.org/0000-0002-7893-7624

Correspondence concerning this article should be addressed to Sara I. McClelland, Department of Women's and Gender Studies, University of Michigan, 204 South State Street, Ann Arbor, MI 48109, United States. Email: saramcc@umich.edu

Psychologists often use measures of group-level prejudice (e.g., racism and sexism) to understand political attitudes and motivations. The Symbolic Racism Scale (SRS; Henry & Sears, 2002) and the Modern Sexism Scale (MSS; Swim et al., 1995) are two especially influential measures of group-level prejudice, as both have been found to predict attitudes toward social issues (e.g., immigration; Berg, 2013), social movements (e.g., the #MeToo movement; Archer & Kam, 2020), and stances toward political candidates (e.g., Maxwell et al., 2013). In measurement development, it is often assumed that the constructs being measured translate directly across groups without consideration of the way that sociopolitical power and historical contexts shape how a group understands and interprets terms researchers use in survey research (McClelland, 2011, 2014). The SRS and MSS were both developed with primarily white[1] samples, and the psychometric equivalence of the measure across race/ethnic and gender groups was not an explicit focus during their development. The present study used measurement invariance analysis to assess the equivalence of the SRS and the MSS in a diverse sample to better understand the extent to which these scales accurately assess racial and gender attitudes toward Black Americans and women.

Literature Review

Marginalized groups are rarely at the center of psychological research in the United States (Goodkind & Deacon, 2004; Henrich et al., 2010), and as a result, little attention is given to how their identities may impact their perceptions and experiences, and how these differences may manifest in scale development (Hall et al., 2016; Hughes & DuMont, 1993). Critical and feminist scholars have recognized that overlooking the influence of lived experiences can be especially harmful to marginalized groups (e.g., ethnic/racial minorities, women; Boag & Tiliopoulos, 2011; Buchanan et al., 2020; Fahs & McClelland, 2016), as neglecting the influence of power and structural forces can lead to inappropriate conclusions about the group (Teo, 2008).

One way to assess how these kinds of assumptions manifest in self-report survey research is to establish measurement invariance (MI). MI determines the psychometric equivalence of a self-report measure across groups, such as those characterized by racial/ethnic identities (Mavondo et al., 2003; Putnick & Bornstein, 2016). If a measure is "invariant," it means that the measure has the same statistical properties across groups. This indicates that the measure effectively assesses the same concept within each group, and that the scores have the same meaning across the groups (see Burnette et al., 2020; Goodwill, 2021; King et al., 2011, for more information). In the present study, we examine the measurement invariance of two commonly used measures: the Symbolic Racism Scale (SRS) and the Modern Sexism Scale (MSS). Two primary research questions guide this study: (a) Is the SRS psychometrically equivalent across racial/ethnic (Black, Latinx, and white) and gender groups (women and men)? and (b) Is the MSS psychometrically equivalent across racial/ethnic (Black, Latinx, and white) and gender groups (women and men)?

Symbolic Racism Scale and Modern Sexism Scale

The SRS (Henry & Sears, 2002) and MSS (Swim et al., 1995) are self-report measures that assess subtle forms of prejudice against Black people and women, respectively. We examined these two scales in the present study because of their shared histories, development, and use in psychological research. They were both developed to study newer, more subtle forms of prejudice in the eras of declining acceptability of overt racism and sexism—Swim et al. (1995), in fact, cited literature on symbolic racism as inspiration for the MSS.

These measures are essential to psychologists who aim to understand political attitudes, as both the SRS and MSS assess prejudice toward groups (i.e., Black people or women), as opposed to other measures of prejudice that focus on assessing prejudice toward individuals. Assessing attitudes about minoritized groups is especially important because they communicate implicit aspects of prejudice, highlighting ways that individuals evaluate a group's characteristics and associated stereotypes. These assessments can be used to effectively predict behaviors (e.g., voting) and endorsement for social policy (e.g., resource distribution or resource restrictions; Albarracín et al., 2018).

Turning to evaluations of racist attitudes, researchers who have sought to examine policy preferences and political attitudes have linked their findings to symbolic racism (Berg, 2013; Ilchi & Frank, 2021; Maxwell et al., 2013; Redlawsk et al., 2014). This form of group prejudice may be so influential because it assesses attitudes toward Black individualism (i.e., the belief that Black people as a group violate American values) and antiequality attitudes (white desire for group-based dominance; Brandt & Reyna, 2012). Thus, symbolic racism may act as a legitimizing ideology, whereby support for policy that inhibits Black progress is justified because Black people are imagined as "less deserving."

The MSS has been used in similar ways to predict attitudes toward social movements (Archer & Kam, 2020), social policy (Martínez et al., 2010), and support for female candidates (Knuckey, 2019). Modern sexism is also thought to act as a legitimizing ideology, as individuals who report high levels of modern sexism tend to deny the existence of discrimination toward women, resent complaints about discrimination, and resent favoritism toward women, all while being against movements and social policy that could disrupt the status quo and improve women's social status (Swim & Cohen, 1997). Overall, the SRS and MSS have important implications for understanding attitudes toward social policy, especially as it relates to policy geared toward marginalized communities, including people of color and women. Thus, ensuring that they are being measured adequately is vital.

Assessing the SRS and MSS

The scales' widespread usage may suggest that they adequately measure racism and sexism across groups; however, for the most part, this has been assumed rather than empirically tested (E.-R. Hayes & Swim, 2013). During the development of the SRS, the authors took care to validate the scale scores with a diverse sample

---

[1] When referring to individuals and samples, we do not capitalize "white" in order to deemphasize the power and normativity that are associated with whiteness that can be communicated with a capital letter. This is consistent with our study's aim to examine the applicability of measures developed using primarily white samples to racially/ethnically marginalized groups. Likewise, when enumerating our groups, we put Black and Latinx before white and women before men to challenge the grammatical norms of listing white people and men first.

so they could ensure that the measure was equivalent across groups (Henry & Sears, 2002). Specifically, they tested the scale among college- and non-college-educated adults, and among white, Black, Latinx, and Asian adults. They found that the measure was internally consistent for the college- and non-college-educated participants, but that internal consistency was low across racial/ethnic groups, particularly for Black and Latinx participants.

Thus, additional work is needed to ensure the validity of SRS scores. Specifically, SRS scores were not validated across gender, so it is unclear whether it is equivalent for women and men. Further, the original study did find that the measure was less reliable for racial/ethnic minorities, so it is uncertain whether the measure was ever conceptually or statistically equivalent for these groups. Second, the original measure was evaluated by examining Cronbach's α, a coefficient that indicates the internal reliability of a measure. However, a measure being internally reliable for multiple groups does not mean that it is conceptually equivalent across groups (Taber, 2018). Additional analyses, such as measurement invariance testing, are needed to determine equivalency (Mavondo et al., 2003).

The MSS is subject to a similar critique. Swim et al. (1995) reported that the scale was unidimensional for both women and men; however, comparisons in internal reliability were not provided, nor was equivalence testing conducted. Further, scale scores were not validated in a diverse group: All 700 of the participants were college students, and nearly all of them were white. Since then, several studies have attempted to reevaluate the MSS, but most used samples that were majority white (Campbell et al., 1997; Morrison et al., 1999). The few studies that did assess racial differences (e.g., E.-R. Hayes & Swim, 2013; Yoder & McDonald, 1997) found reason to question the validity of MSS scores for racial/ethnic minorities. E.-R. Hayes and Swim (2013), for instance, found that the underlying factor structure of the MSS was different in all racial/ethnic groups tested compared to its original validation (one-factor vs. two-factor structure), that white participants were more likely than other groups to endorse the MSS items, and that Black participants were the least likely to endorse items. Further, Cronbach's α was also used in many of these studies as an indicator of conceptual equivalence.

Readers may question why it is important to measure racism and sexism within the groups targeted by such discrimination. Indeed, this belief is likely why there has been limited research assessing the efficacy of the SRS and MSS across race and gender groups. To this end, we note that individuals with marginalized identities are not immune to the social narratives and anti-Black affect that are embedded within U.S. society (Shockley et al., 2016; Smith, 2014). It is possible that Black people hold prejudices and negative attitudes toward their ingroup (Brown et al., 2002), and it is important to assess and explore the prejudices that Black people may have toward their ingroup so that we may address these biases and work toward reducing them.

The Present Study

Given the gaps evidenced in prior validation work, paired with the need to effectively measure racist and sexist attitudes in diverse samples, the present study assessed the equivalence of the SRS and the MSS across race/ethnic and gender groups and identified items that may drive noninvariance. To our knowledge, this is the first study to examine the validity of SRS and MSS scores across both

race and gender and using a measurement invariance analysis. As this is the first foray into measurement invariance testing for these measures, we chose the three largest racial/ethnic groups (Black, Latinx, and white) and two largest gender groups (women and men).

Method

Survey Procedures

In order to examine our research questions by racial/ethnic and gender groups, we used a purposive sample design that targeted potential participants based on their demographics. Participants were recruited via Prolific, an online platform for individuals to find and complete surveys for compensation. Eligibility was contingent upon their answer to two prescreening questions administered. Prolific: their race/ethnicity ("Please indicate your ethnicity [i.e., peoples' ethnicity describes their feeling of belonging and attachment to a distinct group of a larger population that shares their ancestry, color, language, or religion]?") and their gender identity (woman = "female" or "trans female/trans woman" and man = "male" or "trans male/trans man"). Informed consent procedures were completed with participants prior to the start of the survey, and the study was conducted in compliance with the university's internal review board.

Previous research indicated that for an accurate confirmatory factor analysis (CFA), we needed at least 300 participants for each analysis (Putnick & Bornstein, 2016; Worthington & Whittaker, 2006). As a result, we sought to recruit 720 participants using a sampling frame split by participants' racial/ethnic and gender identity, as reported in Prolific's prescreening. We aimed to recruit Black, Latinx, and white participants in equal numbers ($n = 240$), and women and men were sought to comprise equal numbers of each group ($n = 120$).[2] Participants were compensated \$2 ($\sim$\$15/hr) for completing a survey that took approximately 7 min to complete. Study protocols and analysis plans were not preregistered. Data are available via email request from the authors.

Participants

The initial sample included 751 participants. Given our research questions, we removed participants from the data set if they did not identify as a woman or man ($n = 1$, 0.001%), or as Black, Latinx, or white ($n = 7$, 0.009%). Given concerns about the high prevalence of low-quality data collected from online samples (Brühlmann et al., 2020; Chandler et al., 2020; Chmielewski & Kucker, 2020), we implemented many data quality checks to screen out low-quality responses. As a result, we removed four participants (0.005%) for having more than 10% missing data on key variables and 20 participants (0.03%) for failing attention and data checks (see Supplemental Material for more detail).

The final sample was 719 individuals between the ages of 18–75 ($M = 33.6$; $SD = 11.6$). Due to our purposive sampling design, women comprised approximately half of the sample (50.6%; $n = 364$) and the

---

[2] In addition to gender and racial/ethnic diversity, we also recruited for diversity in self-identified political ideology. We recruited for an equal-thirds sample of conservatives, moderates, and liberals and stratified by gender and race/ethnicity, as determined in the Prolific prescreening. However, we did not measure political ideology in our survey, so we do not include a breakdown of participant political ideology in the Participants section.

Black (34.2%; $n = 246$), Latinx (31.0%; $n = 223$), and white (34.8%; $n = 250$) samples were similarly sized. Participants who selected multiple racial and/or ethnic identities including Black were coded as Black, and participants who checked Latinx and other non-Black racial and/or ethnic identities were coded as Latinx. Participants identified as heterosexual (82.8%; $n = 595$), bisexual (11.0%; $n = 79$), lesbian or gay (3.2%; $n = 23$), or another identity (e.g., pansexual, asexual; 2.1%; $n = 15$). Four participants (0.6%) said they did not know how to answer. Three quarters of the sample (75.0%, $n = 539$) reported that they were able to frequently afford essential expenses (e.g., mortgage or rent payment, utility bills, important medical care) in the past 12 months, while 22.5% ($n = 162$) were sometimes able to afford essential expenses, and 2.5% ($n = 18$) were never able to afford essential expenses.

Measures

Symbolic Racism Scale

The eight-item Symbolic Racism 2000 Scale (SRS; Henry & Sears, 2002) was developed to represent the four underlying themes that constitute symbolic racism: (a) the belief that Black people do not work hard enough (e.g., "It's really a matter of some people not trying hard enough; if Black people[3] would only try harder they could be just as well off as white people."); (b) the belief that Black people are demanding too much (e.g., "How much of the racial tension that exists in the United States today do you think Black people are responsible for creating?"); (c) the belief that Black people no longer experience discrimination (e.g., "How much discrimination against Black people do you feel there is in the United States today, limiting their chances to get ahead?"); and (d) the belief that Black people receive an unfair advantage (e.g., "Over the past few years, Black people have gotten more economically than they deserve."). See Supplemental Appendix A for the full scale.

Items were scored on a 4-point Likert scale[4] such that higher mean scores on the scale indicated higher levels of symbolic racist beliefs (after reverse scoring five items). In prior research, the scale demonstrated good internal consistency among white and Asian people ($\alpha > .70$) and poor consistency for Black and Latinx people ($\alpha < .55$; Henry & Sears, 2002). Although Cronbach's α is often used to establish reliability, it is not appropriate for categorical data (Green & Yang, 2009; McNeish, 2018). We instead calculated McDonald's omega (ω), a more general version of α, using A. F. Hayes and Coutts (2020) OMEGA macro for Statistical Package for Social Sciences. McDonald's omega values indicated very good reliability for Black ($\omega = .88$), Latinx ($\omega = .91$), white ($\omega = .92$), women ($\omega = .93$), and men participants ($\omega = .90$).

Modern Sexism Scale

The Modern Sexism Scale (MSS; Swim et al., 1995) is an eight-item scale that taps three underlying dimensions of modern sexism: (a) denial of continuing discrimination (e.g., "It is rare to see women treated in a sexist manner on television."), (b) antagonism toward women's demands (e.g., "It is easy to understand the anger of women's groups in America."), and (c) resentment about special favors for women ("Over the past few years, the government and news media have been showing more concern about the treatment of

women than is warranted by women's actual experiences."). See Supplemental Appendix B for the full scale.

Items were scored on a 4-point Likert scale, with higher mean scores on the scale indicating higher levels of modern sexist beliefs (after reverse scoring five items). Internal consistency of the MSS was good ($\alpha > .70$) in a previous study among Black, Latinx, and white college students (E.-R. Hayes & Swim, 2013). In the present study, McDonald's omega indicated very good reliability for Black ($\omega = .85$), Latinx ($\omega = .89$), white ($\omega = .93$), women ($\omega = .88$), and men participants ($\omega = .88$).

Analysis Plan

Measurement invariance indicates that the same construct was measured in each group, while noninvariance suggests that there may be measurement error for at least one of the groups we examined. We conducted invariance testing using a novel procedure specifically designed for ordered categorical data (Svetina et al., 2020), such as the 4-point Likert-type response options used in this study. Multigroup CFA models were compared on their fit as they became increasingly more constrained. Importantly, this procedure prescribes a revised order for model testing steps that is more appropriate for ordered categorical data (Wu & Estabrook, 2016); these steps are further described in the Results section. We fit four models: SRS across racial/ethnic groups, SRS across gender groups, MSS across racial/ethnic groups, and MSS across gender groups. We used the weighted least square mean and variance adjusted (WLSMV) estimator, which is recommended for ordered categorical data (Garrido et al., 2016). All analyses were completed in Mplus 8.2 (see Supplemental Material for detailed code). Because this altered-order measurement invariance procedure for categorical data is relatively new, psychometricians have not determined ideal incremental fit criteria to our knowledge (D. Svetina, personal communication, February 16, 2022). We thus adopted stringent criteria from the most similar procedures.

We assessed good overall model fit using an RMSEA value $\leq$ .055 per Rutkowski and Svetina's (2017) guidelines for ordered categorical data. Though Rutkowski and Svetina (2017) recommended using a chi-square test of overall fit and not using CFI and TLI for overall fit, we did not adopt this recommendation because the overall sample size was large ($n > 700$), and chi-square tests are overly sensitive in large samples (Putnick & Bornstein, 2016). We did not rely on a chi-square overall fitness test to determine fit (though we report it). Thus, to have multiple fit criteria (as it is recommended to not rely a singular criterion), we also relied on CFI and TLI values greater than .95 per Hu and Bentler (1999) to judge overall fit. We assessed comparative model fit by a chi-square

---

[3] Language in the original scale (Henry & Sears, 2002) referred to "blacks" and "whites" as a group. We changed each instance of "blacks" to "Black people" and "whites" to "white people" per the American Psychological Association's 7th edition language standards. Although item consistency is important, researchers should avoid biased and outdated terminology. As such, we updated the language in an effort to reduce harm to participants. Further, it is not unusual to adjust outdated item phrasing in ways that enhance the clarity of items without affecting the measure's established reliability. See Stevenor and Zickar (2022) for examples and further discussion.

[4] One item on the SRS originally had only three response options. We added a fourth response option to ensure that any differences in variance were not due to scoring differences.

likelihood ratio test for WLSMV paired with a change greater than or equal to .01 in RMSEA and −.004 in CFI, which are recommended for ordered categorical data (Rutkowski & Svetina, 2017).

Results

Configural Invariance

First, we conducted a configural invariance (CI) test to determine the equivalence of the factor structure of the SRS and MSS. To do this, we established a baseline model and tested whether all items had nonzero loadings on the latent factor across all groups. Said another way, configural invariance testing assessed if the factor structure of the scales was the same for each group.

Configural invariance testing showed that a unidimensional factor structure was a good fit for the Modern Sexism Scale across racial/ethnic groups and gender groups, as indicated by RMSEA and CFI (see Table 1). Per modification indices, two pairs of item error covariances that shared themes were allowed to correlate (6 with 7 and 4 with 5) in all groups to improve fit, and we carried forward these covariances in future MI steps.

However, during configural testing of the SRS, we encountered issues with the scale's factor structure. Although the original authors recommend using the Symbolic Racism Scale as a unidimensional scale (Henry & Sears, 2002), we could not establish a unidimensional factor structure for the scale among racial/ethnic or gender groups. The scale authors found that a two-factor structure, in which items were grouped based on whether they were reverse- or forward-scored, better fit the data than a unidimensional factor; the authors interpreted the two-factor structure as a possible artifact of measurement (Henry & Sears, 2002, p. 265). Other researchers testing scales with a mix of reverse- or forward-scoring items have similarly suggested that the item wording can introduce a substantively irrelevant method effect (Marsh, 1996). To investigate this issue further, we mirrored Marsh's (1996) approach and attempted to fit a one-factor model of the SRS across the entire sample (i.e., not divided by race/ethnicity or gender groups) and improve fit by covarying errors between pairs of items that shared the same direction of scoring (either reverse- or forward-scored). We achieved good fit for a one-factor model after four pairs of items' errors were covaried (1 with 2, 1 with 8, 2 with 8, and 6 with 7;

$\chi^2(16) = 32.94, p < .001$, RMSEA = .04, 95% CI [.02, .06]; CFI = .999; TLI = .998), which indicated that the SRS items constituted a unidimensional construct with method effects.

However, the one-factor, covaried-errors model structure on the SRS suffered from a lack of degrees of freedom and was underidentified when specified for configural invariance testing. As a solution, we ultimately tested an alternative two-factor model as suggested by the scale authors (Henry & Sears, 2002; Tarman & Sears, 2005). This approach is further supported by the authors' follow-up study that examined the factor structure of the SRS (Tarman & Sears, 2005). They found that the two factors were highly correlated and that they had near-identical correlations with other relevant constructs (such as support for affirmative action policies and equal opportunity hiring). Moreover, hierarchical regressions revealed little unique predictive power of each factor, that is, the two factors "explain mainly overlapping variance in policy preferences" (p. 747). The authors favored interpreting symbolic racism as one coherent and unified belief system. Future researchers might consider rewriting the SRS to be scored in one direction, as suggested by more recent research (Choi et al., 2022; van Sonderen et al., 2013). Recent research suggests that reverse-scored items do not reduce acquiescent answering or prevent inattentiveness as intended, and they can contaminate participants' scores due to confusion over item meaning. Thus, rewording negatively phrased questions may increase the precision and clarity of the measure. Given that the SRS is a widely used measure and items from the scale will continue to be used in national surveys for the foreseeable future (e.g., the American National Election Studies), we proceeded with the measurement invariance analyses despite the limitations of the scale.

For our analyses, our final configural model grouped forward-scored items (Items 1, 2, and 8) in one factor and the reverse-scored items in a separate, correlated factor (Items 3, 4, 5, 6, and 7). For both the racial/ethnic and gender group models, we allowed the residuals to covary between two items tapping the "excessive demands" theme (6 and 7) in all groups per modification indices, which was theoretically and empirically appropriate. These covariances were freely estimated across groups because we did not pursue strict invariance in this study (Putnick & Bornstein, 2016), and covariances were carried forward in all steps of invariance testing. SRS models in both groups met fit criteria for RMSEA and CFI when modeled this way (see Table 2). We thus

Table 1
Modern Sexism Scale Fit Statistics

Sample and model	$\chi^2(df)$	ΔSig. χ^2	RMSEA [95% CI]	ΔRMSEA	CFI	ΔCFI	TLI	Decision
Women versus men								
Configural invariance	$\chi^2(36) = 67.45, p = .001$	—	0.049 [0.031, 0.067]	—	0.997	—	0.995	Accept
Threshold invariance	$\chi^2(44) = 92.59, p < .001$	***p = .002***	0.055 [0.040, 0.071]	0.006	0.995	−0.002	0.994	Accept
T + loading invariance	$\chi^2(51) = 101.23, p < .001$	***p = .041***	0.052 [0.037, 0.067]	−0.003	0.995	0.000	0.994	Accept
Black versus Latinx versus White								
Configural invariance	$\chi^2(54) = 90.21, p = .002$	—	0.053 [0.033, 0.072]	—	0.997	—	0.996	Accept
Threshold invariance	$\chi^2(70) = 96.24, p = .021$	$p = .923$	0.040 [0.016, 0.058]	−0.013	0.998	0.001	0.998	Accept
T + loading invariance	$\chi^2(84) = 135.50, p < .001$	***p = .004***	0.051 [0.034, 0.066]	**0.011**	0.996	−0.002	0.996	Reject
Partial loading invariance −1 Item	$\chi^2(83) = 105.23, p = .050$	$p = .304$	0.033 [0.000, 0.051]	−0.007	0.998	0.000	0.998	Accept

Note. Bold indicates worsened fit statistic. — = not applicable; *df* = degrees of freedom; Sig. = significant; RMSEA = root-mean-square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker–Lewis index; T = threshold.

Table 2
Symbolic Racism Scale Fit Statistics

| | | | Fit statistic | | | | | |
Sample and model	$\chi^2(df)$	ΔSig. χ^2	RMSEA [95% CI]	ΔRMSEA	CFI	ΔCFI	TLI	Decision
Women versus men								
Configural invariance	$\chi^2(36) = 56.94$, $p = .015$	—	0.040 [0.018, 0.059]	—	0.998	—	0.997	Accept
Threshold invariance	$\chi^2(44) = 66.54$, $p = .016$	$p = .230$	0.038 [0.017, 0.055]	−0.002	0.998	0.000	0.998	Accept
T + loading invariance	$\chi^2(50) = 70.91$, $p = .028$	$p = .269$	0.034 [0.012, 0.051]	−0.004	0.998	0.000	0.998	Accept
Black versus Latinx versus white								
Configural invariance	$\chi^2(54) = 86.93$, $p = .003$	—	0.050 [0.03, 0.069]	—	0.997	—	0.996	Accept
Threshold invariance	$\chi^2(70) = 104.93$, $p = .004$	$p = .218$	0.046 [0.026, 0.063]	−0.004	0.997	0.000	0.996	Accept
T + loading invariance	$\chi^2(82) = 117.71$, $p = .006$	$p = .144$	0.043 [0.023, 0.059]	−0.003	0.997	0.000	0.997	Accept

Note. — = not applicable; *df* = degrees of freedom; Sig. = significant; RMSEA = root-mean-square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker–Lewis index; T = threshold.

established configural invariance among Black, Latinx, and white participants, as well as women and men on the SRS.

Threshold Invariance

The second step to testing measurement invariance with categorical data is to examine the equivalence of item thresholds across groups (Svetina et al., 2020; Wu & Estabrook, 2016). Testing the equivalence of item thresholds is usually tested third (not second), but this order is recommended for ordered categorical data. Item thresholds estimate the values of the latent construct (which is theoretically continuous), where individuals cross over from one ordinal Likert-type response category to the next; item thresholds are conceptually similar to item means when data are continuous. Threshold equivalence testing restricts the levels of item endorsement (i.e., the threshold values) for each item to be the same across groups and compares that fit to the baseline CI model, in which thresholds were allowed to be different across groups. Establishing threshold invariance means that individual differences in the levels of the latent construct accounted for all the differences in the levels of item endorsement. Said another way, individuals with the same latent construct score were more likely to provide similar responses to the items, if the factor loadings were found to be equal across the groups in the third step. The SRS threshold invariance models had good overall fit to the data based on RMSEA and CFI values; the chi-square difference tests were nonsignificant, suggesting that the models' fit was not statistically worse than their respective baseline models; and incremental changes in RMSEA and CFI were within the acceptable limits (see Table 2). The MSS threshold invariance models for racial/ethnic and gender groups similarly did not show worsened fit, as the statistically significant chi-square test in the gender model was not paired with a corresponding RMSEA of CFI fit decrement (see Table 1). As a result, we determined the SRS and MSS to have threshold invariance among Black, Latinx, and white participants, as well as women and men.

Threshold and Loading Invariance

The third step of measurement invariance testing is to determine whether item thresholds and factor loadings are equivalent across groups. Similar to the previous step, the restrictions from the prior step are carried forward (thresholds), and restrictions to factor loadings are added to the model and then compared to the prior model fit (in which the values of the loadings were estimated freely

across groups). When item loadings are invariant across groups, it means that the items reflect the latent construct in a similar way across the groups. Again, this is a revision to the traditional MI order, which constrains loadings, then thresholds.

The threshold and loading models for both race/ethnicity and gender on the SRS and the model for gender groups on the MSS did not have worse fit than their respective threshold models (see Tables 1 and 2). This meant we established threshold and loading invariance for the SRS across racial/ethnic and gender groups, as well as for the MSS across gender groups. However, the MSS threshold and loading model among racial/ethnic groups had worse fit than the previous threshold model. Although overall fit indices were acceptable, the chi-square difference test was significant, indicating the model's incremental fit was significantly worse than the prior model, and the ΔRMSEA increase was greater than .01. These results indicated that at least one item's factor loading was not equivalent across racial/ethnic groups on the MSS.

Partial Loading Invariance on the MSS

To further isolate where there was noninvariance across the racial/ethnic groups in the MSS, we conducted analyses to achieve partial loading invariance. Partial invariance in this case means that one or more factor loadings should be freely estimated while the rest are equivalent (i.e., constrained) across groups. Though there is a lack of consensus about best practices for achieving partial invariance (Putnick & Bornstein, 2016), we tested for partial invariance to provide researchers with clarity as to which items may be problematic and to suggest future research directions, a benefit that outweighed the lack of established best practices. We followed the protocol provided by Svetina et al. (2020), wherein modification indices are used to identify problematic items in the threshold and loading model. Modification indices indicated that Item 8 on the MSS ("Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences") was a potential driver of misfit. We freed the loading for this item across groups, and based on the item loadings within each group, we tested a model in which the loading for this item was freely estimated for white participants but restricted to be equal for Black and Latinx participants. This model had good overall fit to the data, the chi-square test was not significant, and incremental changes in RMSEA and CFI were negligible, indicating that the partial threshold and loading model did not have statistically worse fit

than the threshold model (see Table 1). We therefore concluded that the aforementioned MSS item assessing special favors for women is not invariant among Black, Latinx, and white participants. Figure 1 illustrates the final partial invariance model.

Discussion

In the present study, we sought to understand whether the SRS and the MSS were invariant across racial/ethnic (Black, Latinx, and white) and gender groups (women and men). With regard to the SRS, we were able to establish measurement invariance for all SRS items across race/ethnicity and gender. Our results indicate that researchers may compare SRS scores and their relationship with other constructs between these groups.

Next, we examined the psychometric equivalence of the MSS. We established measurement invariance between women and men on the MSS. We also established configural (i.e., equivalent factor structure) and threshold invariance (i.e., equivalent item thresholds)

Figure 1
Partial Invariance Model of the MSS Across Racial/Ethnic Groups
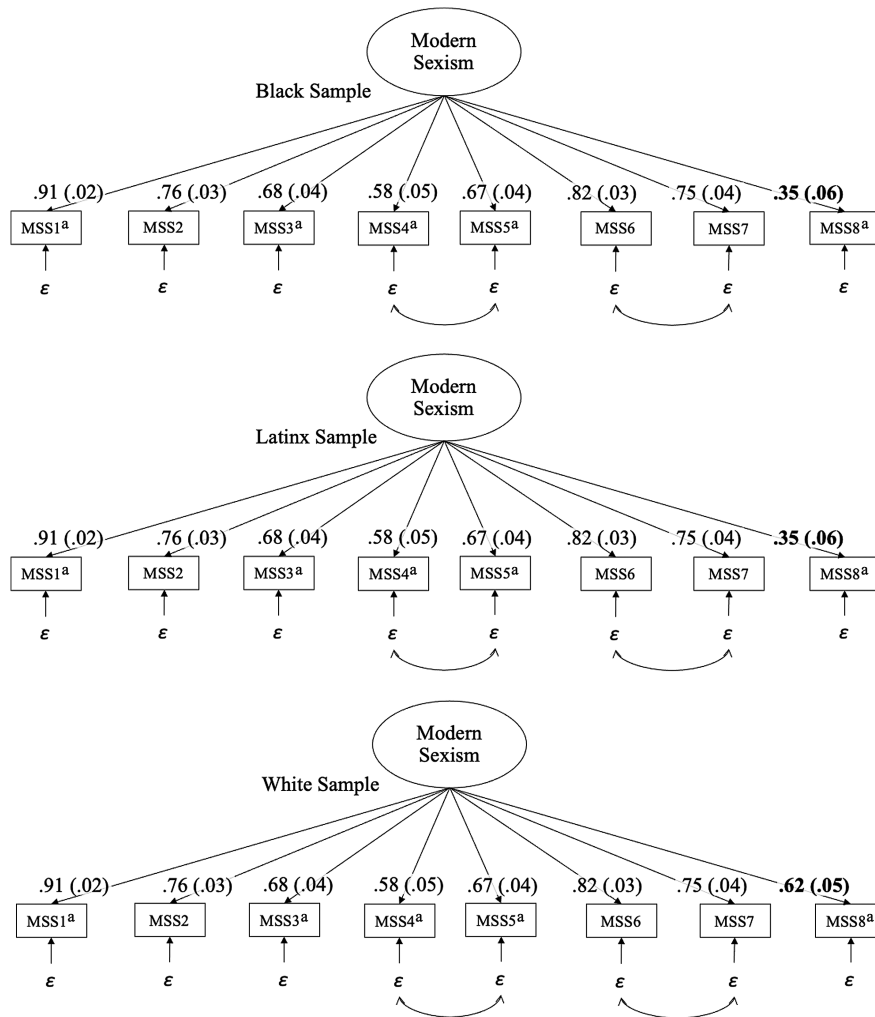
Note. Values next to arrow lines are standardized factor loadings for easy interpretation, standard errors in parentheses. Values in bold (MSS8) were allowed to vary in the white sample versus the Black and Latinx samples during partial threshold and loading invariance testing. Mplus will not provide standardized factor loadings during MI testing; to obtain standardized loadings, we conducted a CFA in Black participants (the referent group for our racial/ethnic group MI analyses). Displayed are standardized loading values from the Black sample in the Latinx and white figures, except for Item 8 (the noninvariant item) in the white figure, whose standardized loading value was obtained from a CFA in the white sample. Thus, this figure is not strictly statistically accurate and is only here to help visualize the findings. Threshold values were omitted for easy presentation, but they were invariant across the three groups for all items. MSS = Modern Sexism Scale; CFA = confirmatory factor analysis; MI = measurement invariance.
[a] Reverse scored items.

across racial/ethnic groups. However, we could not establish threshold and loading invariance for the complete MSS across racial/ethnic groups. In other words, the item loadings were not equivalent.

Consequently, we established partial invariance, which required that the loading of the item, which read, "Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences," be constrained to be equal among Black and Latinx participants but freely estimated for white participants. The MSS is organized around three themes (Swim et al., 1995), and this item represents the theme "resentment about special favors for women." In group-specific CFAs, this item's standardized loading was lower among Black (.35) and Latinx (.38) participants compared to white participants (.62). Further, this item's loading failed to meet both the minimum acceptable value of .40 in exploratory factor analysis for inclusion on a scale and the minimum value of .60 advised for use in CFA. Our findings suggest that the item may not be a meaningful indicator of modern sexism among Black and Latinx people because these groups interpret the item differently than white people. Researchers seeking to use the MSS in diverse racial groups should either reevaluate their use of the problematic item or develop and test alternative ways to equivalently assess the theme the item represents across diverse groups.

When people respond to survey questions, they draw on their own social position and contexts when responding (Davis et al., 2022; McClelland, 2014). One explanation for why the MSS item's loading was noninvariant among racial groups may be due to the lack of concern shown for the treatment of Black women and Latinas by the media and government. Cooky et al. (2010) argued that "media frames are both constructed within raced, classed, and gendered hierarchical relations of power and are read within those very same systems of domination" (p. 144; Slakoff & Fradella, 2019). Slakoff and Brennan (2019) analyzed front-page crime stories about victimized Black, Latina, and white girls and women, and they found that white victims received more sympathetic coverage, whereas coverage of Black and Latina victims invoked racist stereotypes that normalized violence against them. Further, the government has neglected to acknowledge that the treatment of women of color may be both racialized *and* gendered. For example, women of color must meet a higher standard to pursue redress for racialized sexual discrimination because race and sex discrimination have different legal standards (Brassel et al., 2020; Crenshaw, 1990; Leung, 2017). Black, Latinx, and white participants thus may experience media and government concern toward women's issues differently, which could be reflected in the item's contribution to the latent score within each group.

Our results are in line with previous research that has similarly found issues with the MSS as a measure of sexism in people of color (E.-R. Hayes & Swim, 2013) and research finding issues with the item assessing the special favors for women in particular (i.e., "Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences"; Case, 2007). We were able to establish partial invariance in our study; however, the impact of this finding is difficult to assess because we have used a still-developing analytic method. We encourage additional research on

establishing partial invariance and on exploring how issues with single items may or may not affect the implications of the scale. Our findings do not deem the MSS unusable in diverse racial/ethnic groups; however, our results do suggest the need for future researchers examining the factor structure of the MSS to consider how trust in government and media may differ across marginalized groups and how this may impact the interpretation of the problematic item in different populations. It may be beneficial for future researchers to either remove this item or test whether rewording the item may better reflect modern sexist attitudes.

Limitations and Future Directions

This study is important to consider within the context of its limitations. For instance, while we have established noninvariance on the MSS for Black and Latinx participants compared to white participants, we do not know the origins or explanations for that invariance. How are Black and Latinx participants reading and understanding the items compared to white participants? We suggest that researchers gain insight into noninvariance among racial/ethnic groups on the MSS through the use of critical frameworks and methods (e.g., qualitative methods; see Bowleg, 2017, 2021). We specifically recommend employing cognitive debriefing, a method that interrogates how participants understand and respond to in order to gain insight into relations between participants' evaluation of item content, their experience, and their responses (e.g., Rosenbaum & Valsiner, 2011). This method would aid in researchers understanding more about how participants interact with and interpret items that tap difficult to assess constructs, such as racism and sexism. Additionally, we echo the call from others who caution that, given the statistical ambiguity of noninvariance and partial invariance, researchers should attempt to replicate the current findings in independent samples (Putnick & Bornstein, 2016).

For Black and Latinx people, racism plays a role in their understanding of sexism and shapes how they perceive, for example, benevolent forms of sexism (Davis et al., 2022). E.-R. Hayes and Swim (2013) also found that measures of subtle forms of sexism, including the MSS, are not as internally reliable for people of color. The authors suggest that there may be race-based differences in gender-role differentiation and experiences of paternalism. We pose a similar argument: the noninvariance of the item in the MSS regarding media and government concern may stem from Black and Latinx sociohistorical experiences. This important empirical question should be pursued with further research, as it is outside the scope of our analysis of measurement invariance.

Last, we urge readers to refrain from assuming that psychometric equivalence means that anti-Black racism and sexism are conceptualized similarly by ingroup and outgroup members. Kam and Burge (2018) conducted an online, cognitive debrief study of the SRS items and found evidence that the symbolic does, in fact, look different for different racial/ethnic groups. They argued that while conceptualizing the SRS as a manifestation of anti-Black racial resentment among white people is accurate, the measure is actually assessing attributions for inequality (i.e., structural vs. individual) in Black people. Future scholarship should continue to interrogate the meaning of symbolic racism and

modern sexism in diverse groups, including explicitly within groups affected by such attitudes.

Conclusion

This study adds to the growing body of scholarship that seeks to validate in racial/ethnic minority groups the scores of psychological measures originally developed using predominantly white samples (e.g., Burnette et al., 2020; Goodwill, 2021; Granek & Peleg-Sagy, 2017). This work is novel in its examination of the efficacy of the SRS and the MSS across Black, Latinx, and white participants and across women and men. Our findings demonstrate measurement invariance on the SRS (across racial/ethnic and across gender groups) and MSS (across gender groups), providing empirical support for researchers who are interested in the extent to which racist and sexist ideologies are adopted across groups. However, we also found that the problematic item within the MSS, "Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences," was noninvariant among racial/ethnic groups, prompting the need for the item to be reevaluated to improve its statistical contribution to the latent modern sexism score among Black and Latinx participants. We invite continued research into the equivalence of the SRS and MSS across racial/ethnic and gender groups, as refining these measures will provide more precise insight into the endorsement of anti-Black racism and sexism.

References

Albarracín, D., Sunderrajan, A., Lohmann, S., Chan, M. P. S., & Jiang, D. (2018). The psychology of attitudes, motivation, and persuasion. In D. Albarracín & B. T. Johnson (Eds.), *The handbook of attitudes, volume 1: Basic principles* (2nd ed., pp. 3–44). Routledge.

Archer, A. M. N., & Kam, C. D. (2020). Modern sexism in modern times: Public opinion in the #metoo era. *Public Opinion Quarterly*, *84*(4), 813–837. https://doi.org/10.1093/poq/nfaa058

Berg, J. A. (2013). Opposition to pro-immigrant public policy: Symbolic racism and group threat. *Sociological Inquiry*, *83*(1), 1–31. https://doi.org/10.1111/j.1475-682x.2012.00437.x

Boag, S., & Tiliopoulos, N. (2011). *Personality and individual differences: Theory, assessment, and application*. Nova Science Publishers.

Bowleg, L. (2017). Towards a critical health equity research stance: Why epistemology and methodology matter more than qualitative methods. *Health Education & Behavior*, *44*(5), 677–684. https://doi.org/10.1177/1090198117728760

Bowleg, L. (2021). "The master's tools will never dismantle the master's house": Ten critical lessons for black and other health equity researchers of color. *Health Education & Behavior*, *48*(3), 237–249. https://doi.org/10.1177/10901981211007402

Brandt, M. J., & Reyna, C. (2012). The functions of symbolic racism. *Social Justice Research*, *25*(1), 41–60. https://doi.org/10.1007/s11211-012-0146-y

Brassel, S. T., Davis, T. M., Jones, M. K., Miller-Tejada, S., Thorne, K. M., & Areguin, M. A. (2020). The importance of intersectionality for research on the sexual harassment of Black queer women at work. *Translational Issues in Psychological Science*, *6*(4), 383–391. https://doi.org/10.1037/tps0000261

Brown, T. N., Sellers, S. L., & Gomez, J. P. (2002). The relationship between internalization and self-esteem among Black adults. *Sociological Focus*, *35*(1), 55–71. https://doi.org/10.1080/00380237.2002.10571220

Brühlmann, F., Petralito, S., Aeschbach, L. F., & Opwis, K. (2020). The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology*, *2*, Article 100022. https://doi.org/10.1016/j.metip.2020.100022

Buchanan, N. T., Rios, D., & Case, K. A. (2020). Intersectional cultural humility: Aligning critical inquiry with critical praxis in psychology. *Women & Therapy*, *43*(3–4), 235–243. https://doi.org/10.1080/02703149.2020.1729469

Burnette, C. B., Boutté, R. L., Sosnowski, D. W., & Mazzeo, S. E. (2020). Measurement invariance of the SATAQ-4 in Black and White undergraduate women. *Body Image*, *32*, 85–93. https://doi.org/10.1016/j.bodyim.2019.11.004

Campbell, B., Schellenberg, E. G., & Senn, C. Y. (1997). Evaluating measures of contemporary sexism. *Psychology of Women Quarterly*, *21*(1), 89–102. https://doi.org/10.1111/j.1471-6402.1997.tb00102.x

Case, K. A. (2007). Raising male privilege awareness and reducing sexism: An evaluation of diversity courses. *Psychology of Women Quarterly*, *31*(4), 426–435. https://doi.org/10.1111/j.1471-6402.2007.00391.x

Chandler, J., Sisso, I., & Shapiro, D. (2020). Participant carelessness and fraud: Consequences for clinical research and potential solutions. *Journal of Abnormal Psychology*, *129*(1), 49–55. https://doi.org/10.1037/abn0000479

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological & Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Choi, J. W., Thakur, H., Briley, D. A., Temple, J. R., & Cohen, J. R. (2022). Testing the factor structure and measurement invariance of the Temporal Experience of Pleasure Scale in adolescents across time, gender, and race/ethnicity. *Psychological Assessment*, *34*(8), 752–762. https://doi.org/10.1037/pas0001142

Cooky, C., Wachs, F. L., Messner, M., & Dworkin, S. L. (2010). It's not about the game: Don Imus, race, class, gender and sexuality in contemporary media. *Sociology of Sport Journal*, *27*(2), 139–159. https://doi.org/10.1123/ssj.27.2.139

Crenshaw, K. (1990). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*(6), 1241–1299. https://doi.org/10.2307/1229039

Davis, T. M., Settles, I. H., & Jones, M. K. (2022). Standpoints and situatedness: Examining the perception of benevolent sexism in Black and white undergraduate women and men. *Psychology of Women Quarterly*, *46*(1), 8–26. https://doi.org/10.1177/03616843211043108

Fahs, B., & McClelland, S. I. (2016). When sex and power collide: An argument for critical sexuality studies. *Journal of Sex Research*, *53*(4–5), 392–416. https://doi.org/10.1080/00224499.2016.1152454

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods*, *21*(1), 93–111. https://doi.org/10.1037/met0000064

Goodkind, J. R., & Deacon, Z. (2004). Methodological issues in conducting research with refugee women: Principles for recognizing and re-centering the multiply marginalized. *Journal of Community Psychology*, *32*(6), 721–739. https://doi.org/10.1002/jcop.20029

Goodwill, J. R. (2021). Testing for measurement invariance across gender in the 12-item CES-D: An investigation among a nationally representative sample of African Americans. *Cultural Diversity & Ethnic Minority Psychology*, *27*(3), 369–375. https://doi.org/10.1037/cdp0000409

Granek, L., & Peleg-Sagy, T. (2017). The use of pathological grief outcomes in bereavement studies on African Americans. *Transcultural Psychiatry*, *54*(3), 384–399. https://doi.org/10.1177/1363461517708121

Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha.

Psychometrika, *74*(1), 155–167. https://doi.org/10.1007/s11336-008-9099-3

Hall, G. C., Yip, T., & Zárate, M. A. (2016). On becoming multicultural in a monocultural research world: A conceptual approach to studying ethnocultural diversity. *American Psychologist*, *71*(1), 40–51. https://doi.org/10.1037/a0039734

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But … . *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hayes, E.-R., & Swim, J. K. (2013). African, Asian, Latina/o, and European Americans' responses to popular measures of sexist beliefs: Some cautionary notes. *Psychology of Women Quarterly*, *37*(2), 155–166. https://doi.org/10.1177/0361684313480044

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henry, P. J., & Sears, D. O. (2002). The Symbolic Racism 2000 Scale. *Political Psychology*, *23*(2), 253–283. https://doi.org/10.1111/0162-895X.00281

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hughes, D., & DuMont, K. (1993). Using focus groups to facilitate culturally anchored research. *American Journal of Community Psychology*, *21*(6), 775–806. https://doi.org/10.1007/BF00942247

Ilchi, O. S., & Frank, J. (2021). Supporting the message, not the messenger: The correlates of attitudes towards Black Lives Matter. *American Journal of Criminal Justice*, *46*(2), 377–398. https://doi.org/10.1007/s12103-020-09561-1

Kam, C. D., & Burge, C. D. (2018). Uncovering reactions to the Racial Resentment Scale across the racial divide. *The Journal of Politics*, *80*(1), 314–320. https://doi.org/10.1086/693907

King, K. M., Khan, N., Leblanc, P., & Quan, H. (2011). Examining and establishing translational and conceptual equivalence of survey questionnaires for a multi-ethnic, multi-language study. *Journal of Advanced Nursing*, *67*(10), 2267–2274. https://doi.org/10.1111/j.1365-2648.2011.05679.x

Knuckey, J. (2019). "I just don't think she has a presidential look": Sexism and vote choice in the 2016 election. *Social Science Quarterly*, *100*(1), 342–358. https://doi.org/10.1111/ssqu.12547

Leung, K. E. (2017). Microaggressions and sexual harassment: How the severe or pervasive standard fails women of color. *Texas Journal on Civil Liberties & Civil Rights*, *23*(1), 79–102.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*(4), 810–819. https://doi.org/10.1037/0022-3514.70.4.810

Martínez, C., Paterna, C., Roux, P., & Falomir, J. M. (2010). Predicting gender awareness: The relevance of neo-sexism. *Journal of Gender Studies*, *19*(1), 1–12. https://doi.org/10.1080/09589230903057142

Mavondo, F., Gabbott, M., & Tsarenko, Y. (2003). Measurement invariance of marketing instruments: An implication across countries. *Journal of Marketing Management*, *19*(5–6), 523–540. https://doi.org/10.1080/0267257X.2003.9728224

Maxwell, A., Dowe, P. F., & Shields, T. (2013). The next link in the chain reaction: Symbolic racism and Obama's religious affiliation. *Social Science Quarterly*, *94*(2), 321–343. https://doi.org/10.1111/j.1540-6237.2012.00899.x

McClelland, S. I. (2011). Who is the "Self" in self reports of sexual satisfaction? Research and policy implications. *Sexuality Research & Social Policy*, *8*(4), 304–320. https://doi.org/10.1007/s13178-011-0067-9

McClelland, S. I. (2014). "What do you mean when you say that you are sexually satisfied?" A mixed methods study. *Feminism & Psychology*, *24*(1), 74–96. https://doi.org/10.1177/0959353513508392

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Morrison, M. A., Morrison, T. G., Pope, G. A., & Zumbo, B. D. (1999). An investigation of measures of modern and old-fashioned sexism. *Social Indicators Research*, *48*(1), 39–49. https://doi.org/10.1023/A:1006873203349

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Redlawsk, D. P., Tolbert, C. J., & McNeely, N. A. (2014). Symbolic racism and emotional responses to the 2012 presidential candidates. *Political Research Quarterly*, *67*(3), 680–694. https://doi.org/10.1177/1065912914531091

Rosenbaum, P. J., & Valsiner, J. (2011). The un-making of a method: From rating scales to the study of psychological processes. *Theory & Psychology*, *21*(1), 47–65. https://doi.org/10.1177/0959354309352913

Rutkowski, L., & Svetina, D. (2017). Measurement invariance in international surveys: Categorical indicators and fit measure performance. *Applied Measurement in Education*, *30*(1), 39–51. https://doi.org/10.1080/08957347.2016.1243540

Shockley, E., Wynn, A., & Ashburn-Nardo, L. (2016). Dimensions of Black identity predict system justification. *The Journal of Black Psychology*, *42*(2), 103–113. https://doi.org/10.1177/0095798414557276

Slakoff, D. C., & Brennan, P. K. (2019). The differential representation of Latina and Black female victims in front-page news stories: A qualitative document analysis. *Feminist Criminology*, *14*(4), 488–516. https://doi.org/10.1177/1557085117747031

Slakoff, D. C., & Fradella, H. F. (2019). Media messages surrounding missing women and girls: The "missing white woman syndrome" and other factors that influence newsworthiness. *Criminology, Criminal Justice. Law & Society*, *20*(3), 80–102.

Smith, C. W. (2014). Shifting from structural to individual attributions of Black disadvantage: Age, period, and cohort effects on Black explanations of racial disparities. *Journal of Black Studies*, *45*(5), 432–452. https://doi.org/10.1177/0021934714534069

Stevenor, B. A., & Zickar, M. J. (2022). The development and validation of an updated job search behavior scale. *Journal of Vocational Behavior*, *135*, 103722. https://doi.org/10.1016/j.jvb.2022.103722

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using Mplus and the lavaan/semTools packages. *Structural Equation Modeling*, *27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214. https://doi.org/10.1037/0022-3514.68.2.199

Swim, J. K., & Cohen, L. L. (1997). Overt, covert, and subtle sexism: A comparison between the attitudes toward women and modern sexism scales. *Psychology of Women Quarterly*, *21*(1), 103–118. https://doi.org/10.1111/j.1471-6402.1997.tb00103.x

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tarman, C., & Sears, D. O. (2005). The conceptualization and measurement of symbolic racism. *The Journal of Politics*, *67*(3), 731–761. https://doi.org/10.1111/j.1468-2508.2005.00337.x

Teo, T. (2008). From speculation to epistemological violence in psychology: A critical-hermeneutic reconstruction. *Theory & Psychology*, *18*(1), 47–67. https://doi.org/10.1177/0959354307086922

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLOS ONE*, *8*(7), Article e68967. https://doi.org/10.1371/journal.pone.0068967

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*(4), 1014–1045. https://doi.org/10.1007/s11336-016-9506-0

Yoder, J. D., & McDonald, T. W. (1997). The generalizability and construct validity of the modern sexism scale: Some cautionary notes. *Sex Roles*, *36*(9), 655–663. https://doi.org/10.1023/A:1025674209673