"SIS, ARE YOU OKAY?": A FACTOR ANALYSIS TO IDENTIFY THE RAMIFICATIONS OF SYSTEMIC RACISM ON THE PSYCHOLOGICAL WELL-BEING OF BLACK WOMEN IN THE UNITED STATES

Authored By

BENNETT McAULEY
KATELYN SETTLEMYRE
&
DIDIER TURCIOS

North Carolina State University
Department of Statistics

in partial fulfillment of the requirements for
the degree of Master of Statistics;
and in collaboration with our client et al.:

JOCELYN TALIAFERRO, MSW, PhD

North Carolina State University
School of Social Work

&
DIAMOND MEADOWS, MSW

East Tennessee State University
Department of Psychology

May 2023

1 Introduction

The United States has a persistent history of encouraging and enforcing discrimination, oppression, and other forms of prejudice against African Americans, who are often simply characterized as "Black". The turbulent relationship between the Black community and the institutions of government, education, healthcare, law enforcement, social services, and even housing, is a compounded byproduct of historical bylaws–most infamously, the era of Jim Crow laws (c. 1877-1964)–that have yet to be unraveled from the fabric of U.S. American culture and politics.

Racially-motivated violence has been the cornerstone of the Black experience for the entirety of the community's documented history. Paradoxically, Black people are more susceptible to bouts of national trepidation but are expected to show the most resilience. With the additional intersection of gender, Black *women* are especially vulnerable but are often neglected in policies and actions designed to alleviate the stressors of U.S. residents.

In the spring of 2020, the deficit of adequate resources for Black women was proven evident with the surge of COVID-19 and the subsequent deaths of George Floyd and Breonna Taylor. Even with nationwide attention on these issues, Black women continued to suffer in silence under the pretense that they are "strong": perpetually shouldering the burdens of others with no space for their own (Pappas, 2021). Consequently, Black women as a population experience disparities in quality of life that no other group in the U.S. do, most of which can be attributed to social-environmental factors. They are at higher risk for breast cancer, stroke, diabetes, and hypertension; the Black community overall experienced two to three times more complications or death from COVID-19 than the general U.S. population (Menifield & Clark, 2021). Concerning mental health, Black women are half as likely to seek services compared to their white counterparts, and documented studies suggest one reason for this is the lack of cultural competency among providers (Abrams et al., 2018).

In the context of this study, the population of interest is Black women over the age of 18. The barriers previously mentioned have contributed to the marginalization and isolation of Black women in their home lives and in public spaces. By examining potential associations extracted

from respondent data, methods for healthy and impactful coping mechanisms can be identified. Data is derived from a survey conducted in April 2022, to learn how Black women were coping during the pandemic. Approximately 2680 participants responded to 70 questions. In addition to demographic information, the topics covered in the survey included COVID-19, anti-Black racism, depression, anxiety, and coping. The analysis presented in the following sections is driven by structural modeling methods. In lieu of significance testing, the objective of the study is to determine prominent factors that influence the extent to which Black women suffer from depression, anxiety, and other mental health issues.

2   Methodology

2.1   The Data

A survey was conducted via a paid panel to find out how Black women across the United States were coping. Data was collected from April to December 2022, accumulating 2631 respondents. The survey consisted of 70 questions asking about the following domains: COVID-19, Anti-Black Racism, Depression, Anxiety, and Demographic information.

The raw data consists of 157 variables and 2631 subjects. Since this research focuses on Black women, there were questions to screen out participants who were not adult, Black females that consented to the survey. There were 2260 subjects that fit the requirements. Variables that were irrelevant to analysis were also removed, such as personal information or variables that had been recoded, leaving 80 variables for analysis. The data has also been validated by removing potential duplicates (i.e., a respondent had to restart the survey and was recorded twice); and ensuring values for a given variable are all of the same type (e.g., all values for `age` are numeric).

Initially, we believed this was sufficient for implementing the methods outlined in the following sections. During preliminary testing, however, two other issues with the data surfaced: missing values and linear dependencies. Factor analysis methods will not do any calculations with *any* missing values, and often throw an error. They are also susceptible to generating factors with

egregious inflated variances if there are variables that can be expressed as linear combinations of others. So, variables and observations that met the following criteria were filtered out:

- Open-ended responses, as they cannot be applied to any statistical modeling methods
- The proportion of NA observations. Each variable either consists of more than 95% empty observations, or less than 15%. In the former case, the variable was omitted. In the latter case, the observations with the NA value(s) were removed.
- Variables that represent composite scores (the sums or averages of other variables), which were causing the redundancy and linear dependence in the data.

The final dataset consists of 65 variables and 1450 subjects. For many applications, this may be a concern, but as the rest of the report will show, this proved beneficial in containing the volume before applying any algorithmic dimension reduction.

2.2   Structural Equation Modeling (SEM)

Structural Equation Modeling (SEM) is a technique for investigating relationships by specifying a model that represents predictions of that theory among plausible constructs measured with appropriate observed variables (Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007). SEM can be thought of as a combination of regression analysis and factor analysis; it uses *latent* variables or constructs that are measured by multiple observed variables or indicators. For example, an SEM model can be used to estimate the association between outcomes, such as levels of depression or anxiety, and underlying attitudes that influence those, such as coping mechanisms and the effects of Covid-19 or anti-Black racism. SEM allows us to include multiple influences, to allow unobserved concepts that control for the observed indicators, to specify how those concepts influence one another, to assess the model's overall compatibility with the data, and to determine how well the model fits the data (Chapman & Feit, 2019).

To use SEM as an analysis tool, a graphical path diagram of influences is first created. Then, the strengths of the relationships for each path in the model are estimated. The paths in the
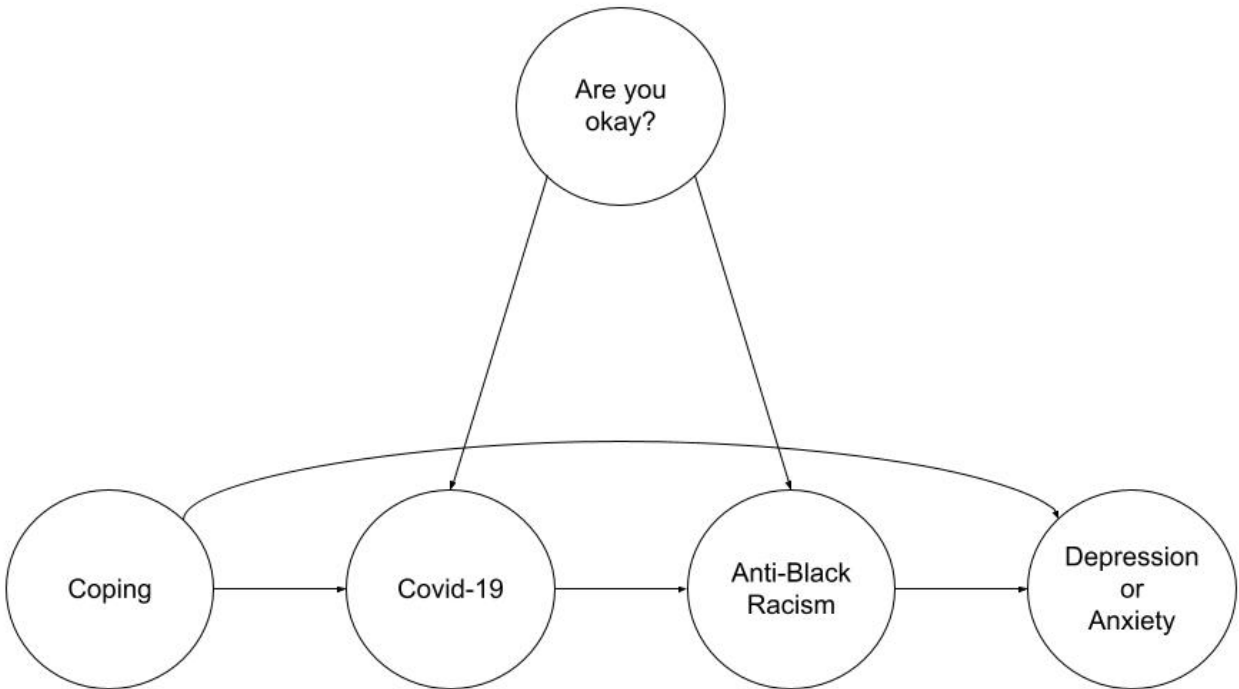
Figure 1: A model measuring levels of depression and anxiety. In this model, coping ability and strategies are associated with Covid-19 and anti-Black racism factors, while the overarching response to *Are you okay?* closely relates to both Covid-19 and anti-Black racism influences, which is then associated with levels of depression and anxiety.

model can be categorized as either observed variables (i.e., have data points), or latent variables that may underlie the observed data. A model is then fitted to the data using the structural model paths created. Figure 1 demonstrates how an SEM model can be constructed to investigate the objectives of this study. For example, the COVID-19 latent variable can be observed as survey items Q1, Q2, and Q3. More complexly, the Coping latent variable can be observed by both survey items *and* other latent variables (i.e., COVID-19, anti-Black racism, etc.). Upon fitting the SEM model, the strength of the relationships between the latent variables are compared, as well as the degree to which the model fits the observed data. Alternative models may then be fit and compared to the original (full) model.

2.3   Exploratory Factor Analysis (EFA)

In Chapter 8.3 of *R For Marketing Research and Analysis*, Chapman and Feit (2019) describe Exploratory Factor Analysis (EFA) as a family of techniques used to assess the relationship of *constructs* (concepts) in surveys and psychological assessments. Factors are regarded as *latent variables* that cannot be observed directly–as one might in a simple linear model–but are associated through their relationship to other variables. For example, abstract concepts such as intelligence, personal preference, and emotions are not observable independently or empirically. Instead, they are measured with a number of indicators known as *manifest variables*, which are empirical factors and behaviors (e.g., GPA, survey responses, and so on) that, in tandem, provide an imperfect explanation to the underlying latent variables.

The goal of EFA is to find the degree to which the latent variables account for the observed variance of the manifest variables. The result is similar to Principle Components Analysis (PCA), where linear combinations or mixtures of the initial variables reduce the dimensionality of the data while retaining as much information as possible. What separates EFA from PCA, however, is the interpretability of the loadings on the variables. In other words, EFA optimizes solutions that favor intuition regarding the original variables so one can test whether the data is consistent with expectations.

To find a solution using EFA, the following steps are generally implemented with a variety of tools and techniques:

1. **Determine the number of factors to estimate and test**. In R, this can be done by constructing a *scree plot*–a line plot of the eigenvalues of factors or principal components–and retaining the factors with an eigenvalue greater than 1.0. An value of 1.0 corresponds to the amount of variance that can be attributed to a *single* independent variable.

2. **Extract factors for modeling**. Determine the variable loadings using PCA, maximum likelihood, or other methods to select the k (limited to the number suggested in Step 1) factors to include and interpret. Trying more than one value of k–none beyond $k \pm 1$–is recommended to determine which model yields stronger interpretability.

3. **Rotate the loadings**. The purpose of rotation is to determine whether correlation between factors should be allowed. Software cannot interpret the context of the *latent variables* but rotation can assist in deciding if, conceptually, factors should be considered independent or if it makes more sense for them to be related. One of two types of rotations can be applied: *orthogonal* and *oblique.* Orthogonal rotation, as its name implies, assumes that factors *do not* correlate. While this method is the default among most software and languages, it is not always the more plausible. Oblique rotation, on the other hand, assumes that factors can and *do* correlate and considers how.

2.4   Why these methods?

The parameters of interest for this study are depression and anxiety. One might opt for regression methods. However, where regression aims to predict a given variable, we are instead interested in the relationship between variables. Both SEM and EFA aim to explain relationships between variables. Additionally, depression and anxiety cannot be measured directly (i.e., they are latent variables), making these methods good options for our data.

One can see the data has high dimensionality. These methods have also been chosen for dimension reduction. Why, then, do we not use principal component analysis (PCA)? PCA *seems* like a good option since it is also a dimension reduction technique. However, the goal of PCA is to

capture as much variability as possible with linear combinations of the variables, called principal components (PCs). It does not make any model assumptions, whereas EFA and SEM propose models and test whether the data fits those models. With PCA, the PCs cannot be interpreted since there is no underlying structure. Our chosen methods have constructs that are easy to interpret.

The two methods we have chosen are very similar, but there is a subtle difference. EFA identifies relationships among variables, whereas SEM is used to determine the extent to which assumptions about the relationships among variables are supported by data (Ockey, 2013). In this study, we identify relationships with EFA and assess these relationships using SEM.

3   Results

3.1   Factor Extraction

The process of extracting factors from our data uses the following R packages: `nFactors`, `GPArotation`, `psych`, `parameters`, `performance`, and `lavaan`.

First, the `check_factorstructure()` function informs us of whether the data is even suitable for factor analysis, based on sampling adequacy and sphericity–whether the correlation matrix of the data is significantly different from an identity matrix.

```
> performance::check_factorstructure(okDat)
# Is the data suitable for Factor Analysis?

  - KMO: The Kaiser, Meyer, Olkin (KMO) measure of sampling adequacy suggests that data seems appropriate for
factor analysis (KMO = 0.92).
  - Sphericity: Bartlett's test of sphericity suggests that there is sufficient significant correlation in the
data for factor analysis (Chisq(2211) = 38924.86, p < .001).
```

Figure 2: Output of preliminary testing for the data. The Kaiser-Meyer-Olkin (KMO) test measure is 0.92. According to Kaiser, values of KMO > 0.9 are marvelous. Bartlett's test of sphericity tests the hypothesis of whether the correlation matrix is an identity matrix. If the null hypothesis is rejected, then we can assume the variables are correlated. Here, the test suggests that the data is significantly correlated and suitable for factor analysis.

From Figure 2, we confirm that factor analysis is the right approach to take for making inference about the data. We expect promising results from this insight, and can move on to

determining how many factors we should attempt to extrapolate. Figure 3 shows the scree plot applied to the data.
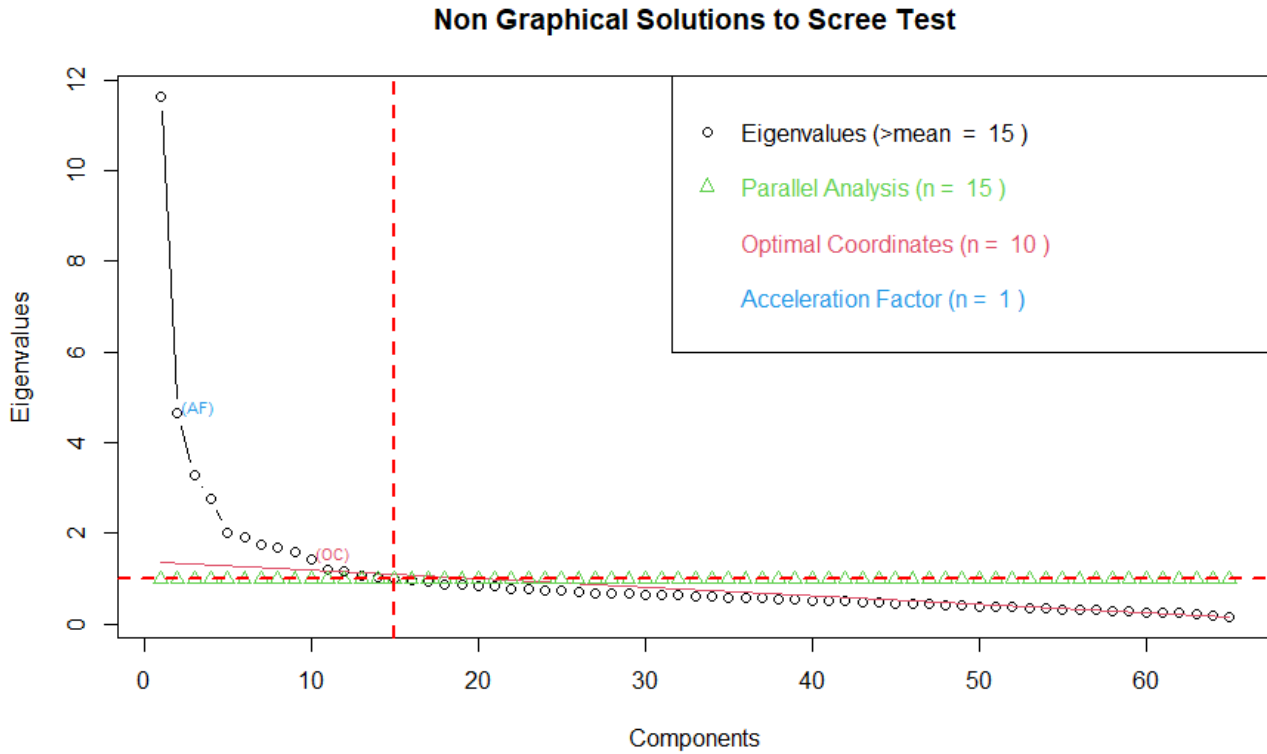

Figure 3: Scree plot applied to survey data. Dashed, red lines at y = 1 and x = 15 indicate the eigenvalue threshold of 1.0 and the largest number of suggested components, respectively. The intersection of these lines is the "optimal" target for factors that should be included.

There is not a unique solution to EFA, but we have two metrics that determined `15` factors is the optimal number to estimate. Considering the number of variables we started with, we believe this is a conservative yet suitable value for interpretability. In this scenario, it does not make sense to test 14 or 16 factors since the adjacent suggested value is 10, and the eigenvalues beyond 15 factors fall below 1.0.

The factor analysis is performed with the `fa()` function, which uses maximum-likelihood for fitting. The following parameter options were supplied:

- `rotation = "oblimin"` - Since is study prioritizes uncovering associations between variables, it is less compelling to assume independence among them, even though orthogonal rotations

are generally simpler to interpret. The difference is also not significant for this data, so `oblimin` is preferred.

- `scores = "regression"` - Thompson's regression method that estimates the unknown element vector of scores f on element vector x to yield $\hat{f} = \Lambda'\Sigma^{-1}x$, where Λ is a matrix of loadings and Σ the correlation matrix of x (Revelle, 2023).

The resulting loadings are listed in the table on page 11 (Figure 4), with our interpretations for each grouping of variables.

3.2 Model Estimation & Evaluation

A multi-stage SEM approach: model specification, model identification, model estimation and evaluation, and model modification was employed in order to evaluate the 15 loadings fitted from the factor analysis. Model specification was comprised of defining relationships among the loadings. Upon confirmation of model identification, subsequent models were evaluated by using a variety of recommended indices such as, chi-square, comparative ft index (CFI), and Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) (Kline, 2010). For the CFI and TLI, a value $\geq .90$ was considered acceptable, and a value $\geq .95$ considered a good fit (Hu & Bentler, 1999). For the RMSEA, a value $< .08$ was considered acceptable, and a value $< .06$ considered good fit (Hu & Bentler, 1999). For the SRMR, a value of $< .05$ was considered a good fit. Although the chi-square statistic are reported, less emphasis is placed on this measure due to its sensitivity to large samples (Hu & Bentler,1999).

The initial hypothesized model, `sem_full`, aimed to include all 15 covariates to evaluate the fit of all the loadings produced from the factor analysis ($\chi^2/\text{df} = 6256.840$, p < 0.001, CFI $= 0.82$, TLI $= 0.81$, RMSEA $= .047$, SRMR $= .074$). Model modification was then conducted to remove non-significant pathways, yielding new and reduced models. Figure 5 and Figure 6 show the final structural models that yielded from model modification, focusing on the outcomes of depression and anxiety, respectively.

Factor #	Variables	Loadings	Description/Interpretation	Factor Name
1	PHQ9_Q1_Little_interest_EST_recode	0.969	*Symptoms of depression from PHQ9 screening questions, with potential for unhealthy coping mechanicms via alcohol/drugs*	phq9_depression
	PHQ9_Q4_tired_low_energy_recode	0.760		
	PHQ9_Q5_Poor_appetite_overeating_recode	0.704		
	PHQ9_Q3_Trouble_sleep_recode	0.696		
	PHQ9_Q6_Feeling_bad_about_self_recode	0.559		
	PHQ9_Q2_Feeling_down_recode	0.557		
	PHQ9_Q7_Trouble_concentrating_recode	0.525		
	PHQ9_Q8_thinking_slowly_fidgety_recode	0.470		
	PHQ9_Q9_SI_Thoughts_recode	0.417		
	PHQ9_Q10_Impact_of_Problems_recode	0.347		
	Coping_alcohol_drugs	0.239		
2	Protests_support_recode	0.740	*Respondant's sentiments regarding racism in the U.S. circa pandemic and subsequent responses to police brutality against Black people*	racism_protest_sentiment
	Protests_sympathy_recode	0.675		
	Protests_appropriate_response_recode	0.561		
	Prior_Covid_Racism	-0.537		
	Followed_Protests_recode	0.471		
	Covid_Racism_increased	-0.446		
	Population_density	-0.156		
3	Coping_nutrition	0.689	*Positive and healthy coping mechanisms; focused on physiological wellbeing*	healthy_coping
	Coping_exercise	0.671		
	Coping_outdoors	0.648		
	Coping_pleasurable_activities	0.535		
	Coping_spiritual	0.502		
4	Covid_wellbeing_Emotional_recode	0.863	*Impact of COVID-19 pandemic on aspects of respondent's wellbeing*	pandemic_impact
	Covid_wellbeing_psychological_recode	0.810		
	Covid_wellbeing_overall_recode	0.768		
	Covid_Wellbeing_Physical_recode	0.725		
	Covid_wellbeing_financial_recode	0.590		
	Covid_wellbeing_social_recode	0.584		
5	Number_adults_household	0.816	*Respondant's household size*	household_size
	Number_household	0.836		
6	Prior_Covid_experienced_racial_discrimination_recode	0.835	*Respondant's experience of racial discrimination circa pandemic*	exp_racism
	Covid_experienced_racial_discrimination_recode	0.915		
7	Academia_status	0.484	*Aspects of respondent's personal identity and mindspace, and what of that may have brought them to take the survey*	personal_id
	Protest_Participation	0.317		
	Political_views	0.312		
	Sexual_Orientation	-0.274		
	Find_out_study	0.257		
	Age	0.109		
8	Covid_dx	0.807	*Respondent's COVID-19 vaccination and isolation status*	covid_dx_status
	Covid_isolate	0.527		
9	Income	0.660	*Respondent's socioeconomic status/level*	socioecon_status
	Education	0.515		
	Marital_status	-0.292		
10	Approval_Trump_Protests_recode	0.635	*Respondent's sentiments regarding how the government handled the police brutality protests*	govt_sentiment
	Protests_police_good_job_recode	0.624		
	Protests_propertydamanage_undermine_recode	0.364		
11	GAD7_Q3_worrying_too_much_recode	0.897	*Symptions of anxiety (GAD) from GAD7 screening questions*	gad7_anxiety
	GAD7_Q2_cannot_stop_worrying_recode	0.867		
	GAD7_Q7_afraid_something_bad_recode	0.735		
	GAD7_Q4_trouble_relaxing_recode	0.717		
	GAD7_Q1_Feeling_nervous_Recode	0.663		
	GAD7_Q5_restless_recode	0.566		
	GAD7_Q6_irritable_recode	0.554		
12	Coping_movies_TV_activities	0.510	*Coping mechanisms to target emotional resiliancy; focused on psyhological wellbeing*	emotional_coping
	Coping_distract_work	0.471		
	Coping_optimism	0.388		
	Coping_humor	0.358		
	Coping_not_show_emotions	0.320		
	Coping_emotional_support	0.305		
13	Vote_2020	0.603	*Respondent's decision making habits influenced by their political ideology*	politics_descisions
	Political_affliation	0.552		
	Covid_Vaccine	0.240		
14	Covid_death	0.546	*COVID-19 diagnosis and prognosis outcomes for people in respondant's life*	covid_impact_others
	Covid_family_Friends	0.495		
15	Employment_status	-0.443	*Respondant's employement status*	employment_status

Figure 4: Table of the variables, loadings, and the factors that will serve as the basis for the SEM model. Descriptions and names for each factor are provided. The variables highlighted in purple are below the moderate correlation threshold of 0.2, and provide no significant contribution to the factor of which they are assigned.

The structural equation model for depression, `sem_depression`, presented a relatively adequate fit (χ^2/df = 1166.054, p < 0.001, CFI = 0.94, TLI = 0.93, RMSEA = .051, SRMR = .051). This model included five covariates: `Exp_Racism`, `Pandemic_Impact`, `Restless`, `Depression_Shame`, and `Psychological_Coping`. There was a significant main effect of the impact of the COVID-19 pandemic on aspects of respondent's well-being, and the symptoms of anxiety and depression related to restlessness and unhealthy or out-of-character means of managing it (β = 0.54, p < 0.001). There was also a significant effect on a respondent's experience of racial discrimination circa pandemic, and symptoms of anxiety and depression related to restlessness (β = 0.25, p < 0.001). All five covariates in this model were significantly associated (p < 0.001).

The structural equation model for anxiety, `sem_anxiety`, presented an adequate fit (χ^2/df = 498.206, p < 0.001, CFI = 0.95, TLI = 0.94, RMSEA = .041, SRMR = .047). This model also included five covariates:`Political_Decisions`,`Covid_Impact`,`Exp_Racism`,`Anxiety_Uneasy`, and `Psychological_Coping`. There was a significant negative main effect of a respondent's COVID-19 diagnosis and prognosis outcomes in their life (for themselves and others), and a respondent's symptoms of anxiety related to feeling uneasy or tense (β = −0.21, p < 0.001). Additionally, there was a significant effect on a respondent's symptoms of anxiety, and coping mechanisms to target emotional resiliency and let the mind rest, focused on psychological well-being (β = 0.45, p < 0.001). All covariates in this model were significantly associated, with the exception of the association between `Political_Decisions` and `Covid_Impact` (β = 0.03, p < 0.434), as well as the association between `Political_Decisions` and `Exp_Racism` (β = 0.02, p < 0.518).

4 Discussion

The current study investigated the associations between various factors surrounding the COVID-19 pandemic and anti-Black racism that influence the extent to which Black women suffer from depression, anxiety, and other mental health issues. Our results demonstrated significant associations between (1) symptoms of depression related to feeling shame or failure and difficulties with executive functioning, (2) coping mechanisms to target emotional resiliency and psychological
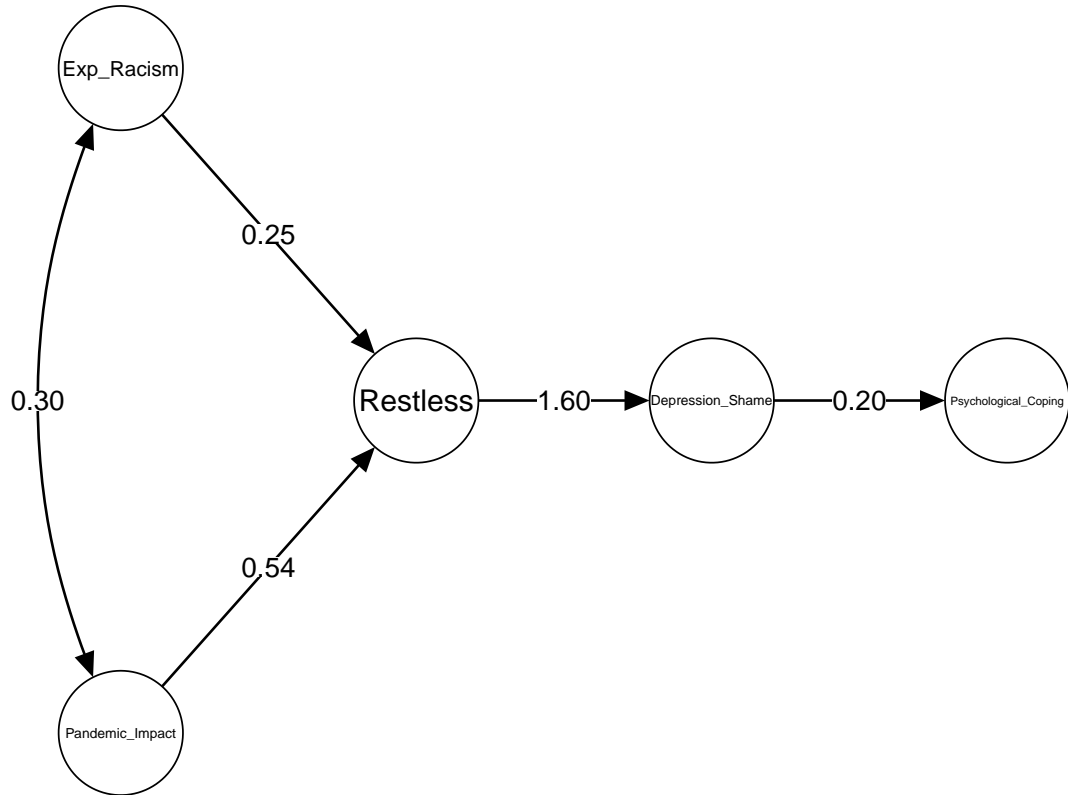
Figure 5: SEM_Depression. Structural Equation Model with Standardized Loadings for Depression. N = 1,450. Depression_Shame: Symptoms of depression related to feeling shame/failure and (or because of) difficulties with executive functioning. Psychological_Coping: Coping mechanisms to target emotional resiliancy and let the mind rest; focused on psychological well-being. Pandemic_Impact: Impact of COVID-19 pandemic on aspects of respondent's well-being. Exp_Racism: Respondant's experience of racial discrimination circa pandemic. Restless: Symptoms of anxiety and depression related to restlessness and unhealthy/out-of-character means of managing it
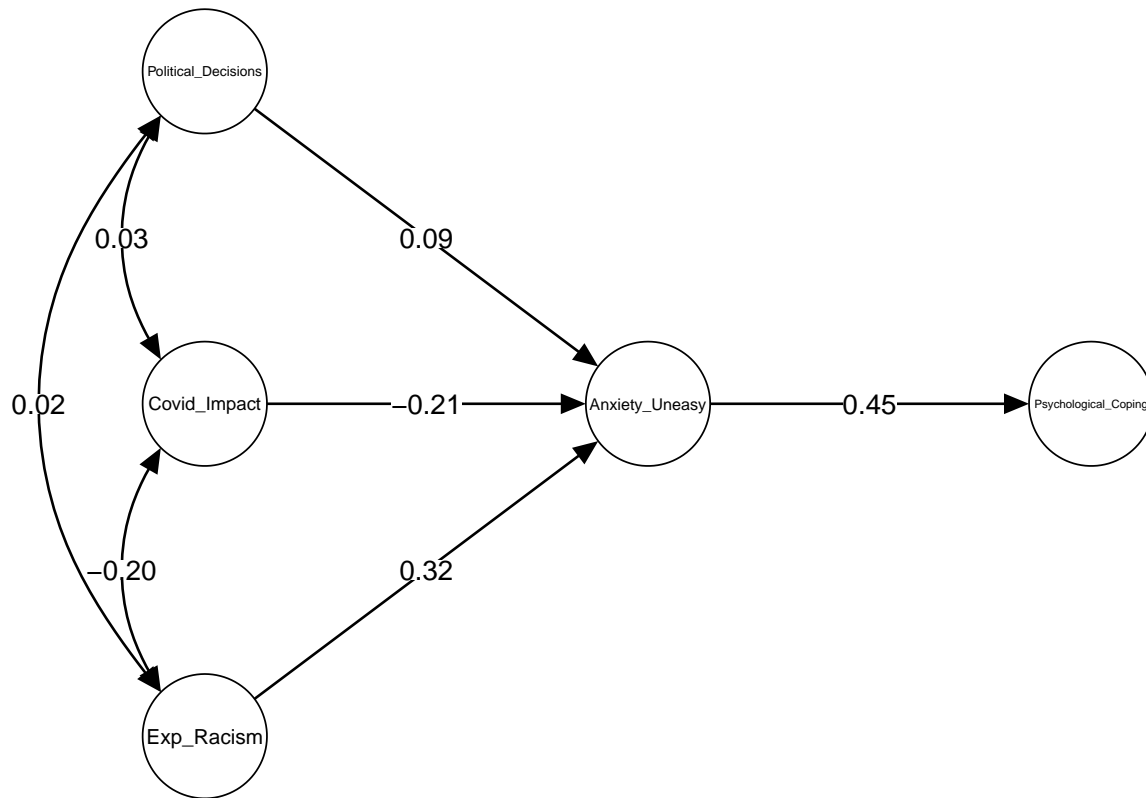
Figure 6: SEM_Anxiety. Structural Equation Model with Standardized Loadings for Anxiety. N = 1,450. Anxiety_Uneasy: Symptoms of anxiety related to feeling uneasy/tense. Psychological_Coping: Coping mechanisms to target emotional resiliancy and let the mind rest; focused on psychological well-being.Exp_Racism: Respondant's experience of racial discrimination circa pandemic. Covid_Impact: COVID-19 diagnosis and prognosis outcomes in respondant's life, for themselves and others. Political_Decisions: Respondant's decision making habits influenced by their political ideology.

well-being, (3) the impact of the COVID-19 pandemic on aspects of a respondent's well-being, (4) a respondent's experience of racial discrimination circa pandemic, and (5) symptoms of anxiety and depression related to restlessness and unhealthy or out-of-character means of managing it.

Additionally, our results indicated a significant negative association between a respondent's COVID-19 diagnosis and prognosis outcomes in their life (for themselves and others), and a respondent's symptoms of anxiety related to feeling uneasy or tense. Symptoms of anxiety were also indicated to be significantly associated with (1) coping mechanisms to target emotional resiliency and psychological well-being, (2) a respondent's experience of racial discrimination circa pandemic, and (3) a respondent's decision making habits influenced by their political ideology.

References

Abrams, J. et al. 2018. *Underneath the Mask of the Strong Black Woman Schema: Disentangling Influences of Strength and Self-Silencing on Depressive Symptoms Among u.s. Black Women.* Sex Roles, 80(9-10), 517–526. https://doi.org/10.1007/s11199-018-0956-y.

Chapman, E., C. & Feit. 2019. *R for Marketing Research and Analysis, Second Edition.* Springer Nature Switzerland AG.

Hayduk, Cummings, L. 2007. *Testing! Testing! One, Two, Three—Testing the Theory in Structural Equation Models! Personality and Individual Differences.* Personality; Individual Differences, 42, 841–850.

Hu, & Bentler, L. T. 1999. *Cutof Criteria for Ft Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives.* Structural Equation Modeling, 6, 1–55. https://doi.org/10.1080/10705519909540118.

Kline, R. B. 2010. *Principles and Practice of Structural Equation Modeling.* Guilford Press.

Matsuzaka, Avery, S. 2022. *Online Victimization, Womanism, and Body Esteem Among Young Black Women: A Structural Equation Modeling Approach.* Sex Roles 86, 681–694. https://doi.org/10.1007/s11199-022-01296-z.

Menifield, & Clark, C. E. 2021. *Pandemic Planning in the United States: An Examination of COVID-19 Data.* Public Administration Review, 81(6), 1102–1109. https://doi.org/10.1111/puar.13326.

National Center for Education Statistics. 2022. *Characteristics of Postsecondary Faculty. Condition of Education.* U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/csc.

Ockey, G. J. 2013. *Exploratory Factor Analysis and Structural Equation Modeling.* In The Companion to Language Assessment, A.J. Kunnan (Ed.). https://doi.org/10.1002/9781118411360.wbcla114.

Pappas, S. 2021. *Effective Therapy with Black Women.* American Psychological Association. https://www.apa.org/monitor/2021/11/ce-therapy-black-women.

Revelle, W. 2023. *Package 'Psych': Procedures for Psychological, Psychometric, and Personality*

Research. CRAN. https://cran.r-project.org/web/packages/psych/psych.pdf.