# DEVELOPMENT AND VALIDATION OF A COMPREHENSIVE STEREOTYPICALITY MEASURE

Kyjeila Latimer, M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2022

APPROVED:

Yolanda Niemann, Major Professor
Anthony Ryals, Committee Member
Martinque Jones, Committee Member
Donald Dougherty, Chair of the Department of Psychology
James Meernik, Interim Dean of the College of Liberal Arts and Social Sciences
Victor Prybutok, Dean of the Toulouse Graduate School

Latimer, Kyjeila. *Development and Validation of a Comprehensive Stereotypicality Measure.* Doctor of Philosophy (Behavioral Science), August 2022, 36 pp., 1 table, 1 appendix, references, 63 titles.

Racial stereotypicality refers to the degree to which an individual looks like a "typical" member of their ethnic or racial group by considering multiple phenotypical features such as skin tone and nose width. Prior studies have utilized real and photoshopped images to assess perceptions of individuals high in racial stereotypicality. However, no known studies have allowed participants to engage in the self-assessment of their own facial features outside of skin-tone. In the present study, I develop and investigate the underlying structure of a scale which allows Black individuals to self-assess their perceived degree of racial stereotypicality. I accomplished this by developing items, soliciting expert feedback, conducting cognitive interviews, disseminating the proposed scale, and conducting an exploratory factor analysis (EFA) on a sample of 308 Black adults. EFA results produced a three-factor structure influenced by item wording and reverse coding. Findings also indicated that items which assessed one's overall degree of stereotypicality loaded onto a singular, separate factor as originally theorized. Results suggest that reverse coding, item wording, and response labeling may influence factor structure and negatively impact scale validation procedures. Additionally, items assessing overall stereotypicality may address something distinctly different from other items which assess individual features. Therefore, perceived overall racial stereotypicality should be further tested and considered in future research since it performed fairly during exploratory analysis, aligns with proposed theory, and ultimately homes in on perceptions that may have major implications for understanding how Black phenotypical features impact the lives, outcomes, and experiences of Black individuals.

TABLE OF CONTENTS

Page

# INTRODUCTION

Skin tone is a salient, visual cue that leads to implicit and explicit stereotype activation (Foy & Ray, 2019; Maddox & Gray, 2002). Stereotypes are a function of cognitive processes we utilize to organize and explain our surrounding world (Adams et al., 2016; Maddox & Gray, 2002). Within prior literature, the focus has typically been on the role of skin tone as a salient indicator of racial group membership (Branigan et al., 2017). Skin tone has been utilized to explore ingroup and outgroup dynamics via research on stereotyping, discrimination, and bias, among other constructs (Allport, 1954; Brewer, 1999; Dixon & Maddox, 2005; Richeson & Sommers, 2016). However, recent literature has focused on how differences in skin tone effect both intergroup and intragroup racial dynamics via skin tone bias.

Skin tone bias, also known as colorism, color consciousness, or being color struck describes the differential treatment typically experienced by individuals within a racial group based off of the lightness or darkness of their skin (Breland, 1988; Harvey et al., 2017; Russel, Wilson, & Hall, 2013). For the purpose of this review, I will use the term "skin tone bias," as it encompasses both the stratification of individuals by skin tone and multiple forms of bias that may occur as a function of said stratification (Adams et al. 2016). Researchers have studied how lighter skin, being more adjacent to European, Eurocentric norms has historically been associated with beauty, privilege, refinement, and greater wealth (Dhillon-Jamerson, 2018; Harris, 2018l Hill, 2002), and resulted in preferential treatment such as greater media representation (Adams et al., 2016; Breland-Noble, 2013); socioeconomic status, occupational achievement, educational attainment (Alon et al., 2019; Ryabov, 2013); and skin tone satisfaction (Alon et al., 2019).

Scholars have also highlighted how, as a result of colorism, darker skin is often discriminated against and associated with negative stereotypes such as criminality,

unattractiveness, unintelligence, poverty, hypermasculinity, and aggression (Blake et al., 2017; Hairston et al., 2018). Furthermore, numerous studies have found that individuals with darker skin are more likely to: be racially profiled and experience discrimination (Blake et al., 2017; Maddox & Gray, 2002); have lower self-esteem (Adams et al., 2016; Landor et al., 2019; Thompson & Keith, 2001); have poorer health outcomes (Landor & Smith, 2019; Sweet et al., 2007); report experiencing psychological duress (Maddox & Perry, 2018), experience depression (Maddox & Perry, 2018), and are at greater risk of receiving school suspension and disciplinary action at school (Blake et al., 2017; Hannon et al., 2013). These studies, focused on gradations in skin tone, have allowed for a more nuanced look at differences among members of the same ethnic minority groups. However, skin tone is not the only racial phenotypic characteristic that has been researched and utilized to determine physical difference within ethnic/ racial groups.

Racial stereotypicality has also received attention in the extant literature and refers to the degree to which an individual looks like a "typical" member of their ethnic or racial group[1] (Hebl et al., 2012). Racial stereotypicality more thoroughly explains varying phenotypical features beyond skin-tone that make individuals look more or less like a stereotypical member of their racial group, by considering additional features like hair color, nose broadness, and eye size (Hebl et al., 2012). Understandably, indicators of racial phenotypicality differ by racial group. For example, stereotypical phenotypic features for Black individuals include thicker lips, wider noses, tightly coiled, as well as darker hair and skin, while stereotypical phenotypic features for Asian individuals include smaller, almond shaped eyes, darker hair, and fuller cheeks (Caldwell, 2003; Hebl et al., 2012; Johnson & Bankhead, 2014). Similar to skin-tone, these additional

[1]Race is a social construct which refers to the classification of people by genotype and phenotypical features (Burton et al., 2010). Ethnicity refers to groups who share a common culture, ancestry, religion, or nationality that separates them from other groups and can occur across or within racial groups (Daniel, 2002).

secondary racial characteristics have been linked to a number of outcomes. Research indicates that greater perceived stereotypicality influences outcomes related to the perceived (individuals with highly stereotypical features) and outside perceivers and effects may vary by racial group. Among Black individuals, greater racial stereotypicality has been associated with an increased likelihood of experiencing social rejection online (Hebl et al., 2012); greater anxiety over police treatment and harassment (Kahn et al., 2017); an increased likelihood of being discriminated against (Kahn et al., 2017); and a greater likelihood of receiving the death penalty (Eberhardt et al., 2006). Findings also suggest that perceivers are more likely to view highly stereotypical Black faces as threatening (Kleider-Offutt et al., 2018). Research by Kahn and Davies (2010) found that during the classic shoot/don't shoot task, perceivers were more likely to exhibit both implicit stereotyping and explicit bias toward highly stereotypical faces with results indicating that stimuli with these features were more likely to be wrongfully shot.

Blair and colleagues (2004) found that perceivers negatively stereotyped both Black and White faces with more Afrocentric features. Furthermore, even when perceivers were explicitly informed of the cognitive processes underlying their stereotyping, they were unable to avoid the process of automatically assigning them to highly Afrocentric faces (Blair et al., 2004). However, for White individuals, highly stereotypical Eurocentric features serve as a protective mechanism and are associated with decreased use of severe police force and greater academic motivation (Kahn et al., 2016; Williams et al., 2019). Greater racial stereotypicality among Asian individuals has been associated with low academic motivation and a greater likelihood of experiencing psychological distress (Lee & Thai, 2015; Williams et al., 2019). Outside perceivers are more likely to assume that Asians high in stereotypicality have greater academic ability in the field of STEM and are less attractive than those lower in stereotypicality (Wilkins,

Chan, & Kaiser, 2011; Williams et al., 2019). Lastly, research on stereotypicality among Latinx individuals suggests that greater stereotypicality predicts poorer academic outcomes (Ryabov & Goza, 2014).

Though the outcomes for non-white individuals with darker skin-tone and more stereotypical features are similar, prior research suggests that facial features and skin tone do not equally influence perceptions and in fact demonstrate independent, additive effects on both explicit and implicit outcomes (e.g., measures of liking, typicality ratings) (Hagiwara et al., 2012; Stepanova & Strube, 2009). For this reason, I argue that a scale which explicitly assesses both skin-tone and other physical dimensions of stereotypicality needs to be conceptualized and validated through further research within this domain. The purpose of this study is to create and validate an objective, self-assessment of stereotypicality that is inclusive of both facial features and skin tone for Black individuals. Though the effects of colorism have been studied internationally, work on racial stereotypicality has only been observed domestically and primarily focused on Black individuals (Dixon & Telles, 2017; Sims & Hirudayaraj, 2016). For this reason, the current project will only focus on Black individuals residing within the United States. The following section will overview prior methodological approaches to assessing the effects of, and perceptions toward, various racial phenotypic features.

## History of Related Measures in the Extant Literature

### Digital Manipulation

In attempts to better understand the effects of stereotypical features and gradations in skin tone on the perceptions of observers, a number of researchers have relied on technological software like Adobe photoshop. For instance, Opie and Phillips (2015) consulted with a professional expert to photoshop stock images of professional White women with similar,

Eurocentric straight hairstyles along with stock images of Black women donning either Eurocentric, straight hairstyles or Afrocentric, natural hairstyles (afro hairstyle or dreadlocked hairstyle). Participants were randomly exposed to one photoshopped image and asked to assess professional appearance, likelihood of success, and perceived dominance (Opie & Phillips, 2015). [Results indicate that evaluators perceive Black female applicants less professionally and less likely to succeed when they don Afrocentric hairstyles (Opie & Phillips, 2015). Results also indicated that Black evaluators judged applicants with Afrocentric styles more harshly than White applicants and this relationship was mediated by perceptions of the applicant as overtly dominating (Opie & Phillips, 2015).

Similarly, Cowart and Lehnert (2018) digitally manipulated the skin-tone of White, Black, and Hispanic models on stock photos obtained from an online provider. First, model photos were faced forward, dressed similarly, and shared a similar facial expression. Next, the photos deemed highest in stereotypicality by undergraduate participants for male and female models across three racial groups (Black, Hispanic, and White) were selected for use in the final phase of the study. Photos were then altered by a professional graphic designer so that each photo had a lighter and darker skinned condition. Finally, one written item was explicitly utilized to assess the skin-tone of the stimuli on a Likert scale ranging from 1 (“very light for that ethnic group”) to 7 (“very dark for that ethnic group”). The skin tone manipulation for White models was not significant and cross race analyses indicated that neither Black nor White models received lower evaluative ratings as a function of skin tone. Therefore, the researchers chose to only focus on within group variations among Hispanic models and results pertaining to the effect of skin-tone for the Black and White models were not presented. Contrary to colorism/skin tone bias theory, the white-collar lighter-skinned Hispanic model was seen as less competent and

evaluated more poorly than the white-collar Hispanic model presented with darker skin-tone (Cowart & Lehnert, 2018). This effect was primarily driven by participants who self-identified as having darker skin and may suggest that some advantages of lighter skin tone may be determined by whether perceivers are more proximal to Whiteness themselves (Cowart & Lehnert, 2018).

Similarly, Harrison and Thomas (2009) utilized Adobe Photoshop to digitally alter photos so that they appeared to have light, medium, or dark-hued skin and findings revealed that light and medium toned applicants were perceived as more competent and given higher recommendations than their darker toned counterparts. As part of their methodology, Hairston and colleagues (2018) also digitally manipulated the face of a photographed Black male to present as having either light, light-medium, medium-dark, or dark skin to subtly determine whether participants' perceptions of a stimuli's mental health varied as a function of skin-tone. However, results indicated no significant differences in counseling student's perceptions of mental health by target's skin-tone.

Lastly, Hagiwara and colleagues (2012) utilized digital manipulation during a four-step process to assess effects of both skin tone and phenotypicality by first looking at over 100 photos of Black men with neutral facial expressions who were posed in the same manner, excluding photos that were racially ambiguous. During this stage, the lip thickness, face height, face width, width of the nose, and the ratios of lip thickness to face length as well as nose width to face width were calculated. The skin tone (as assessed by luminosity of pixels) for each face was also measured. Part two of this process entailed selecting the 20 darkest-skinned and 20 lightest-skinned target photos and digitally manipulating nose and lip ratios (by one SD above or below the mean), so that four conditions of photos were created: darker skin with less prototypical features, darker skin with more prototypical features, lighter skin with less prototypical features,

and lighter skin with more prototypical features. During step three, a pilot study was conducted to assess the effectiveness of facial feature manipulation alone. This was achieved by digitally altering each face to have no skin tone which was achieved by editing the faces into black/white silhouettes with no shading. Participants were then asked to assess the stereotypicality of each face. In the final step, a total of 64 photos were selected, half of which possessed less stereotypical facial features, while the other half possessed more stereotypical facial features. Furthermore, each of the facial feature conditions were further divided so that one half of the target photos had dark skin and the other half had light skin, resulting in a total of four conditions. Lastly, an implicit, sequential priming task was utilized to assess emotional reactions along with an explicit Likert-type measure to assess how much participants liked target faces. The results from this study revealed that White individuals had greater implicit and explicit negative reactions toward individuals with darker skin-tone and more stereotypical features (Hagiwara et al., 2012). Findings also suggest that though the effects of skin-tone and facial features are additive, the effects of skin-tone have a greater impact on negative perceptions toward Black individuals (Hagiwara et al., 2012). Research utilizing digital manipulation techniques has generally been used to observe how outside perceivers evaluate facial stimuli with varying skin tones and degree of stereotypicality. The next section will focus on interviewer ratings, which utilize outside perceptions to assess and code features of a target group members.

### Interviewer Ratings

Independent raters have also been historically utilized to assess phenotypical features, specifically skin-tone. In the seminal Doll study by Clark and Clark (1947) interviewers assessed children's skin-tone and categorized them into one of three nominal categories (light, medium, or

dark) in order to assess how skin tone effected both racial preferences and racial identification[2]. Alon and colleagues (2019) utilized three interviewers to assess participant skin-tone and establish interrater reliability. All interviewers in this study based their skin-tone assessments on the forearms of the participants, justified by the fact that this part of the body is less likely to darken and change color over time (Alon et al., 2019). All interviewers originally rated skin-tone on a 5-point scale with values which included: light, medium-light, medium, medium-dark, and dark. All three interviewer scores were then averaged and condensed into three categories (light, medium, or dark) by score (Alon et al., 2019). Findings from this study indicated that darker skinned gay males were more likely to have partners with higher educational attainment and more likely to view race as an integral part of their male identity. Williams and colleagues (2019) incorporated interviewer assessments of stereotypicality into their protocol by utilizing participant photos and asking interviewers to rate the degree to which individuals looked like typical members of their racial group on a scale of 1 (not at all typical) to 5 (very typical). Black women with greater stereotypical features were perceived as having worse STEM ability. Participants who were not highly concerned about appearing prejudiced also associated Black men with highly stereotypical features with poorer STEM skills (Alon et al., 2019). In the following section, methodological approaches that rely directly on the self-assessments of target group members, is detailed.

### Self-Perceived Assessments

Studies have also allowed respondents to self-report their own perceived skin-tone and

[2] Researchers recently replicated the classic doll study to assess children's skin tone preference and self-concept (Byrd et al., 2017). Unfortunately, the researchers did not detail methodology for how the skin tone of participants were assessed. However, results indicated that children attributed positive attributes to the light-toned doll instead of the White or black doll which illustrated both racial awareness and a potential skin tone preference for fairer skin. Children participants also perceived the dark-toned doll as having a "mean" attitude and showed a preference for dolls with long, straight hair over short or Afrocentric hairstyles (Byrd et al., 2017).

degree of phenotypicality. An early study by Hamm and colleagues (1973) allowed respondents to indicate their own skin tone by choosing from an array of 11 identical faces, ranging in color from white to dark brown. Results from the study indicated that participants were realistic in the self-selection of their skin-tone. Similarly, in order to assess participant's perception of their own stereotypicality, Kahn and colleagues (2017) distributed a one item questionnaire which stated: "Other people think I physically look like a typical member of my racial/ethnic group" and allowed participants to indicate the degree to which they agreed with the statement on a Likert type scale. Findings from this study revealed that individuals who were high in perceived stereotypicality were more likely to express concern over receiving poor police treatment as a function of their race. Several other studies have allowed individuals to self-report their own-skin tone and features alongside other methods like interviewer ratings. The following section will further elaborate on studies which have utilized mixed approaches to assessing features related to phenotypical stereotypicality.

## Mixed Methodology

Past studies have combined different methodologies to assess varying aspects of phenotypicality and skin-tone. For example, Averhart and Bigler (1997) assessed self-perceived skin-tone by utilizing five tiles that varied from light tan to dark brown and asked children to select the colored tile that was closest to their own skin tone. In addition to self-assessments, two interviewers of the same race also independently rated the children's skin-tone on the same scale. Results indicated that averaged interviewer ratings and children's own self-assessed ratings were similar and deviated no more than one standard deviation for over 75% of the child respondents. Interviewers also demonstrated reliability agreeing on 93% of all cases (Averhart & Bigler, 1997). Interestingly enough, in the cases where rater discrepancy was demonstrated, children

were more likely to assign themselves a skin-tone rating much lighter than that assigned by either interviewer (Averhart & Bigler, 1997). Similarly, Coard and colleagues (2001) utilized a mix of both self-assessments and interviewer ratings to obtain skin tone data. Self-assessments were obtained by asking participants to pick the shade closest to their own facial skin-tone from the Skin Color Assessment Procedure (SCAP), a 9-point color palette ranging from "light cream" to "dark ebony". Respondents were also given a three-item Skin Color Questionnaire (SCQ) which included one item that specifically assessed self-perceived skin color on a scale of 1 (*extremely light*) to 9 (*extremely dark*). In addition to self-assessments of skin-tone, interviewers also covertly assessed participant's skin-tone on a 9-point scale. Satisfactory interrater reliability was demonstrated between interviewers' skin tone judgements by a reliability coefficient of .90 and the average interviewer ratings also significantly correlated with both the self-assessed SCAP ratings ($r$ = .72) and the self-assessment questionnaire item from the SCQ ($r$ = .68). Lastly the color palette self-assessment (SCAP) and item from the questionnaire self-assessment (SCQ) were also significantly correlated ($r$ = .66). Deficits and limitations of the aforementioned measures will be detailed in the following section.

## Limitations of the Extant Literature on Stereotypicality

Though researchers have attempted to assess both skin tone and facial features as markers of stereotypicality, there are key issues with past methodological approaches to these assessments. One key limitation of past studies are inconsistencies in interviewer-based assessment methodologies (Blake et al., 2017). For example, the Massey-Martin scale is an established scale that has been widely utilized by interviewers to assess skin-tone (Massey & Martin, 2003). Because the Massey-Martin scale's protocol does not allow for a direct comparison of skin tone palettes to skin, researchers are trained to memorize participants' skin

tone and code at a later time. The drawback of this methodology is that the lapse between engaging with participants and coding forces interviewers to rely on their memories of participants, which has the potential to introduce bias or error (i.e., misinformation) into a study's design. Hannon and DeFina's (2016) investigation of the Massey-Martin scale's reliability, found discrepancy between interviewers' skin tone ratings given in 2012 and 2014, evidenced by low intra class correlations (> .5). Furthermore, their findings also suggest that racial mismatch may lead to skewed skin-tone assessments as White interviewers are less able to detect skin tone variability and are more likely to assign darker skin tone ratings to participants than Black interviewers (Hannon & DeFina, 2016).

Additionally, though the 10-point, Massey-Martin scale has been widely utilized, it is not the only range of tones that have been utilized in former studies—with other palettes for both interviewer and self-assessed skin-tone ranging anywhere from 5 to 11 points (Blake et al., 2017; Brown et al., 1999; Dixon, 2019; Hannon & DeFina, 2016; Hersch, 2018). The lack of a standardized point system for skin-tone palettes makes it difficult to objectively define what we classify as light skin versus medium-tone or darker skin—a distinction which could be particularly helpful in legal cases or research focused on the discriminatory/ detrimental effects of skin-tone bias. Past studies have also utilized differing reference points for assessing skin-tone, with some studies recommending that the tone of forearms (Alon et al., 2019) be utilized while others urging that facial color be utilized (Massey & Martin, 2003) for assessments, which does not allow for standardized comparisons across studies. A final glaring limitation of past studies is that several studies have allowed for the self-assessment of skin-tone but not considered creating a tool that assesses other phenotypical features of participants along a gradient of increasing stereotypicality. Specifically, past studies have utilized images and

photoshopped stimuli to assess perceptions of highly stereotypical facial features (e.g., Hagiwara et al, 2012; Kahn & Davies, 2010), but none have allowed participants to engage in self-assessment of their own facial features to gauge how attributes such as fuller lips and kinkier hair contribute to their lived experiences, life outcomes (e.g., self-esteem), or experiences with discrimination.

## Filling a Void in the Literature

I suggest that it is time to begin filling the voids in the extant literature. Specifically, the development and validation of a comprehensive measure that includes skin tone as well other salient, physical phenotypic features would allow for a formal, standardized account of stereotypical, physical features that could be utilized across studies. For the purpose of this study, relevant phenotypic features include hair texture, hair color, eye color, nose width, and lip thickness, as these have been cited as markers of stereotypicality by prior authors and in the present study I only aim to identify physical features that are readily visible to perceivers. Other physical attributes beyond facial features will not be included as part of the scale, primarily because stereotypical characterizations of Black bodies differ by gender with Black men being stereotyped as taller and more muscular/athletically built and Black women being stereotyped as being heavier/curvier (Kwate & Threadcraft, 2015; Wilson et al., 2017). The proposed scale would be the first to allow individuals to self-assess both skin tone and other visual phenotypical features such as nose width, lip thickness, and hair texture.

## Significance

Mixed method studies such as that of Coard and colleagues (2001), which properly incorporate both in-person interviewer assessments and self-assessments, find that these two methods are significantly correlated, suggesting that participants can accurately self-report their

own skin-tone. Likewise, research indicates that self-assessed perceptions of feature based stereotypicality are highly related to interviewer assessments. The development of a self-assessment measure of overall stereotypicality that does not rely on interviewer ratings may offer significant advantages. Such a measure that does not require interviewer assessment allows for remote assessment of phenotypical features which is particularly advantageous during the current, COVID-19 pandemic. Additionally, this measure could be easily implemented in qualitative or quantitative studies which are generally not-in person (e.g., qualitative Zoom interview, Qualtrics studies) or in settings where interviewer assessments may be limited by lighting, accessibility, or bias due to interviewer-participant racial mismatch.

Though Hagiwara and colleagues (2012) found that physical features of stereotypicality function independently of skin tone, few studies have considered the importance of independently parsing out physical, phenotypic features from skin tone. The proposed Perceived Black Stereotypicality scale (PBS) would allow for each dimension of stereotypicality to be separately weighted to determine how various parts which define stereotypicality ultimately influence individuals' perceptions, experiences, and behavior.

Furthermore, the scale would allow for easy self-assessment of features by using a mixture of Likert-type questions and pictorial guides that participants can utilize for direct comparisons. Lastly, this scale allows for physical evaluations of Black individuals which does not simply homogenize them as a racial group, and allows for more detailed, in-depth analyses of within- group differences.

## METHOD

A formal scale development study typically occurs in two stages and involves cognitive testing, two rounds of survey administration, exploratory factor analysis (EFA), convergent and divergent validity checks, inter-item correlation assessments and a confirmatory factor analysis (CFA). However, this study only aimed to develop the BSS scale, perform cognitive pretests, modify items, and assess the underlying factor structure of the proposed scale by conducting an EFA. The proposed measure of Black stereotypicality was formulated and tested in two major phases: (1) item generation and (2) scale development (Boateng et al, 2018). Phase 1 focused on deductively identifying domains and dimensions, generating items, and incorporating expert evaluations. The proposed Perceived Black Stereotypicality Scale (PBS) is an objective self-assessment (including pictorial and Likert-type items) of stereotypicality that is inclusive of facial features, and skin tone for Black individuals. The initially generated items relied on participants' self-assessment of visual, and facial features which have been noted in prior literature as indicators of stereotypicality and or Blackness (nose width, lip thickness, skin-tone, hair color, eye color, and hair texture) (Caldwell, 2003; Hagiwara et al., 2012; Hebl et al., 2012; Johnson & Bankhead, 2014), along with self-perceptions of overall perceived stereotypicality. A total of five dimensions were originally theorized to encapsulate these physical features: (a) *physiognomic features* focuses on the size of features on a person's face that have been historically associated with racial typicality (e.g. nose and lips), (b) *hair texture* focuses on how kinky/curly one's natural hair is, (c) *feature colors* which focuses on hair and eye color, (d) *skin* tone focuses on the lightness or darkness of one's skin, and (e) *overall stereotypicality* which centers on how individuals self-assess their own perceived degree of stereotypicality.

Proposed self-assessment items for phenotypic attributes were presented to participants

with five images of a feature which ordinally increased in perceived stereotypicality from left to right along with Likert type questionnaires that allowed individuals to self-assess their degree of stereotypicality along multiple dimensions. All presented features were gender neutral, so as not to force a false gender binary. For example, to assess hair texture, participants were asked to pick from an array of five photos which will vary in kinkiness from a chemically processed, straight hair follicle to a tightly coiled hair follicle. Because these pictorial self-assessment items are unipolar, a five-point scale was utilized as recommended by Boateng and colleagues (2018). Additionally, experts suggest that the initial pool of generated items be twice as long as the intended scale (Schinka et al., 2012). Because a 5-dimension scale is being proposed, a minimum of 40 items were generated for the initial item pool.

Next, I consulted with academic professionals who have expertise in colorism, race, and body image along with professionals who have experience with scale development. Each expert was given a scale assessment tool that allowed them to classify each item into dimensions and evaluate each item's degree of fit. The assessment tool also included an open-ended portion for additional feedback (E.g., What additional features, if any, are a central part of Black individual's perceived stereotypicality?") Feedback from these experts were utilized to inform the acceptance, modification, or rejection of items.

Phase 2 of the study focused on scale development and validation. During the first stage, cognitive pretesting of Black community members was utilized to assess face validity. Face validity is the extent to which members of the target population determine that items appropriately reflect the constructs that they intend to (Hayes et al., 1995).

Boateng and colleagues (2018) suggest conducting between 5 to 15 interviews in two to three rounds or until no new insights emerge during interviews. Therefore, five Black college

students and five Black individuals from the surrounding community were recruited to pre-test items. A mixture of online flyers and snowball sampling were utilized to recruit Black interviewees from these two groups. All participants were invited to a two-part structured interview. During the first portion participants were asked to read aloud and answer each item. Interviewers asked occasional follow-up questions for clarification and to understand respondents' comprehension of terms and concepts. During the second portion of cognitive pretesting, participants were asked to discuss what physical features accounted for Black stereotypicality. All feedback was utilized to further revise items for survey dissemination.

After all interview data was collected, the actual survey was administered to Black participants via a survey link on both Amazon's Mturk and Facebook. Because the original iteration of this scale is intended for Black individuals, participants were required to identify as Black before participating in pretesting.

# RESULTS

## Preliminary Expert Feedback

Quantitative expert feedback was utilized to assess which items indicated poor fit, by using descriptive statistics to average judge scores. Items with low scores were eliminated along with items given a score of zero by multiple raters. As a result, 3 items were eliminated. All expert quantitative data on dimensions fit was gathered and input into Excel. Majority vote was utilized to decide what dimensions items best fit into. All items with majority vote classifications which clashed with the original, intended classification were removed. Mismatch indicated that items did not clearly capture the proposed dimension. Three additional items were deleted as a result. All items that were given no classification at all by the expert panel were also marked but had already been marked for deletion, as they were given the lowest rankings by the expert panel.

Qualitative expert feedback was used alongside quantitative scores to further revise items and to understand why items were rated poorly or favorably. Qualitative comments from multiple judges indicated that Items 33 and 34 were too similarly worded to items that had been previously displayed. Therefore, both were eliminated leaving a total of 32 items.

## Preliminary Community Sample Feedback

A total of 10 Black respondents participated in cognitive pretesting. Results from pretesting indicated that participants struggled to select pictorial images that closely resembled their own features without referring to their Zoom webcams. The hardest image to properly match was hair texture, because participants were unsure of their own textures or claimed to have a mix of hair textures. Participants who struggled with pictorial images of facial features felt that they were between sizes and had to approximate by sizing up or down to the nearest

representative photo. Most participants affirmed theorized features related to stereotypicality and listed skin tone, hair texture, nose and lips, as features that helped in identifying potential Black racial group membership. Additionally, participants believed that certain features, notably skin tone played a greater role in determining stereotypicality than other features like hair and eye color. All participants understood "stereotypicality" as a spectrum, with higher stereotypicality relating to more Afrocentric features. Some respondents questioned whether bodily features were also meant to be considered when thinking through questions. In response, "facial" features were specified whenever possible to provide clarity.

## Survey Administration and Data Cleaning

The 32-items and a demographic survey were advertised to participants on Amazon's mTurk and on Facebook. In order to participate, respondents had to identify as Black and be over the age of 18 years old. All respondents who did not identify as Black during the eligibility prescreening, were immediately rerouted to the end of the survey and thanked for their participation. A total of 999 cases met prescreening requirements and were recorded in Qualtrics. All reverse-coded items were properly transformed to prepare for data cleaning and analysis. Twenty-one cases were deleted for missing more than 5% of data. 155 cases were deleted for completing the survey in under 2 minutes. 18 cases were deleted for not passing validation checks. 105 cases were deleted for providing nonsense answers in open-response demographic text boxes (e.g. entering "Wakanda" as birth country). Remaining cases were flagged as candidates for deletion if they completed the survey too quickly (under 240 seconds), had duplicate IP addresses, displayed straightlining behavior[3] or racially identified with 4 or more

[3] Straightlining takes place when respondents provide the same or similar answers to a group of questions, which may negatively impact data quality (Kim et al., 2018). Straightlining behavior may be indicative of respondents

racial groups. All remaining cases with two or more flags were deleted from the dataset, leaving a total of 308 cases after data cleaning.

## Participants

The final, clean sample consisted of 308 U.S. citizens who racially self-identified as Black. Of the 308 participants, 41% identified as male, 59% identified as female, and less than 1% identified as queer. Additionally, 73% of participants identified as straight/ heterosexual, 2% identified as gay or lesbian, 21% identified as bisexual, 2% identified as questioning, less than 2% identified as pansexual, asexual or queer. Participants ranged in age from 19 to 87 ($M$ = 38.5, $SD$ = 11.4).

## Assumptions and Normality

All data was screened for outliers by utilizing Mahalanobis distance testing and 12 items were identified as outliers. Of the 308 remaining cases, 35 were missing data. Little's MCAR test was conducted and indicated that data was missing completely at random ($\chi 2 = 70.08$, $df = 60$, $p = .175$). Because each case was missing less than 4% of data, mean imputation was utilized to estimate responses for cases with missing responses. Univariate skewness and kurtosis were assessed for all variables and did not exceed suggested limits. However, Shapiro-Wilks tests suggest that the data is not normally distributed ($df = 308$, $p < .001$).

Cronbach's alpha was utilized to assess the reliability of each theorized dimension and the overall scale. Items within each dimension which held low average correlations with other items were dropped to raise the alpha level to a minimum acceptable level of .70, which is recommended for conducting EFA (Watkins, 2018). Dimension one was originally composed of

racing through survey responses without comprehending items or not attempting to provide real answers or opinions to items.

8 items and produced a Cronbach's alpha of .776. Item 1.1 was removed, increasing the alpha value to .79 . Dimension two was originally composed of 8 items and produced a Cronbach's alpha of .73 . Dimension three was originally composed of 5 items and produced a Cronbach's alpha of .79 . Item 3.1 was removed, increasing the alpha value to .80. Dimension four was originally composed of 5 items and produced a Cronbach's alpha of .50. Items 4.4 and 4.5 was removed, increasing the alpha value to .69 . Dimension five was originally composed of 6 items and produced a Cronbach's alpha of .65 . Item 5.4 was removed, increasing the alpha value to .76. The overall alpha level of all items (excluding Items 1.1, 4.4, 4.5, and 5.4) was .83 . Next the factorability of all items were evaluated. The Kaiser-Meyer-Olkin measure of sampling adequacy was .88, exceeding the recommended value of .6, and Bartlett's test of sphericity was also significant ($\chi 2$ (351) = 3867.62, $p < .05$). Finally, almost all item communalities were above .30, further confirming that each item shared some common variance with other items. Given these overall indicators, factor analysis was deemed to be suitable with the remaining 27 items.

## Factor Analysis

The statistical software package utilized to conduct the EFA was version 28 of SPSS. Principal axis factor analysis was used because the primary purpose was to identify latent structure of proposed items and because it is suitable for the development of new scales. Additionally, principal axis factoring methods are optimal for data which is not normally distributed. The first initial, unrotated EFA was conducted. Six factors were identified according to the "eigenvalue greater than 1 rule", and approximately five factors were suggested according to the scree plot yielded. Initial eigenvalues indicated that the first two factors explained 26% and 17% of the variance respectively. The third, fourth, fifth and sixth factors had eigenvalues just over 1, and explained 5%, 5%, 4%, and 4% of the variance, respectively. Ultimately five

factors were decided upon as suggested by the scree plot and the original theory of how items should be clustered. A second EFA was run after extracting five factors and utilizing promax rotation and accounted for 49.38% of variance. However, cross-loadings on 4 items became evident. To attain simple structure and drop items with communalities below 0.3, Items 4.1, 4.3, and 2.3 were deleted. These modifications left only two items on Factor 5 and three items on Factor 4, which suggests that fewer factors needed to be considered. The third EFA was conducted with only four factors and accounted for 50.23% of variance. Only two factors loaded onto the fourth factor, suggesting fewer factors were still needed. The fourth and EFA was conducted by extracting three factors and utilizing promax rotation. To eliminate cross-loading, Items 3.2, 1.5, 1.3, 3.1, 2.1, 5.1, and 4.1 were deleted for cross-loadings above 0.3, low communalities below 0.3, or low factor loadings below 0.5. For the final stage, a principal axis factor analysis of the remaining 17 items, using promax rotations, was conducted, with three factors explaining 53.2% of the variance. The factor loading matrix for this final, three-factor solution is presented in Table 1. The coefficient alphas for each factor were moderate: .64 (95% CI = .57-.70) for the first factor, .74 (95% CI = .69- .79) for the second factor, and .72 (95% CI = .66-.76) for the third factor. Factors 2 and 3 correlated at .59, Factor 1 and 2 correlated at -.233, and Factor 1 and 3 correlated at -.178. Given these results, the three-factor solution was accepted as an adequate structural representation of the PBS measure.

In summary, the prior analyses suggest a 3-factor structure. Factor 1 was labeled Reverse Coded because though it included items which assessed varying facial physical features (e.g., nose width, lip fullness, hair color, eye color, and hair texture), all of the reverse coded items loaded onto this factor. Specifically, the items which loaded the highest onto this factor were all reverse coded and items on this factor which were *not* reverse coded correlated negatively with

all other items on the factor. Factor 2 was labeled Kinky Texture because it only included items which strictly assessed the kinkiness of one's hair texture. All Likert labels for items on this factor rated hair by texture (*very loose/ straight* to *very kinky/ tightly coiled*) and did not include items with Likert labels that assessed degree of agreement (*strongly disagree* to *strongly agree*). Factor 3 was labeled Overall Stereotypicality because it solely included items which assessed overall stereotypicality and did not include items that assessed individual features.

Table 1

*Factor Loadings and Communalities based on Principal Axis Factor Analysis with Promax Rotation for 17 items of the Perceived Black Stereotypicality (PBS) Scale (N = 308)*

| | Factor Loading | Commun-ality |
|---|---|---|
| Factor 1: "Reverse Coding" (α = .64, % variance = 32.2) | | |
| 1.4) "Compared to most other people within the Black community, I have a broad/ wise nose." | -.51 | .36 |
| 1.6) Other Black people believe that the width/size of my nose is: | -.51 | .42 |
| 1.7) Other Black people believe that the fullness/size of my lips is: | -.34 | .38 |
| 1.8) Compared to most other Black people, I believe that the fullness/size of my lips is: | -.46 | .45 |
| 2.2) "Other Black people think I have very loose or straight hair." | .88 | .74 |
| 2.5) "I have very loose/ straight hair." | .81 | .62 |
| 2.6) "When other Black people look at the texture of my natural hair, they question my racial group membership." | .84 | .69 |
| 3.3) "Other Black people think that my eye color is not typical for members of my racial group." | .78 | .61 |
| 3.4) "Compared to most people within the Black community, the color of my eyes is different." | .84 | .69 |
| 3.5) "Compared to most people within the Black community, the color of my hair is different." | .87 | .76 |
| Factor 2: "Kinky Texture" (α = .74, % variance = 16.3) | | |
| 2.4) Other Black people would agree that my hair texture is composed of kinky, tight coils." | .53 | .33 |

*(table continues)*

| | Factor Loading | Commun-ality |
|---|---|---|
| 2.7) “Compared to most people within the Black community, I believe that the natural coils in my hair are:” | .96 | .75 |
| 2.8) “My natural hair texture is:” | .71 | .48 |
| Factor 3: “Overall Stereotypicality” ($\alpha$ = .72, % variance = 4.7) | | |
| 5.2) “As a Black individual, I believe that I physically look like a typical member of my racial group.” | .71 | .40 |
| 5.3) “As a Black individual, I have many physical, facial features that are typical of members of my racial group (e.g., tightly coiled hair, full lips, broad nose).” | .55 | .40 |
| 5.5) Using your own perception of how you think a “stereotypically Black” face looks, how stereotypically Black do you think your face is? | .65 | .50 |
| 5.6) How “stereotypically Black” do other Black individuals think you physically look? | .65 | .44 |

*Note*. Factor loadings < .3 are suppressed.

## DISCUSSION

The current paper provides a foundation for developing a measure of Black phenotypic stereotypicality and assessing how various features correlate. Findings from the study suggest that dimensions within a measure of stereotypicality are not distinguished by feature, but rather how items are presented. Question wording and framing significantly impacted how individuals rated and perceived their own features. For example, respondents answered questions differently when considering how other people within the Black community view and assess features versus when they reflected on how they perceive their own features. As a result, Factor 1 consisted almost entirely of items that referenced the perceptions of "other Black people" and included items which asked about feature tones, physiognomic features, and texture.

Items that utilized reverse coding also impacted factor structure. Specifically, reverse coded items which utilized polar opposite wording loaded onto the same factor and were negatively correlated with all other "reverse coded" items that did not utilize reverse coding. This might suggest that participants may not interpret the phrasing of words as intended or not see them as contradictory statements. For example, a participant may firmly agree with the following reverse coded statement "Compared to most people within the Black community, the color of my hair is different" and select five on a Likert scale. However, they might also select a pictorial option for dark brown hair which would be coded as a four on the same scale. Technically, both responses would be fairly accurate considering that the participant's hair tone is not the same color as most Black individuals, yet it is still fairly dark in hue. This issue might suggest that the reverse coding of items warped the intended meaning and validity of some proposed items. The loading pattern of all reverse coded items onto a singular factor also suggests that EFA is sensitive to the methods used. These findings align with work by Zhang and

colleagues (2016) who found that the number and type of reverse coded items significantly impacted factor structure and created a “method” factor when performing EFA.

Similarly, “Kinky Texture” was differentiated by the way in which items were worded. All items were unique in that they specifically asked participants to indicate how kinky/ curly their hair texture is. “Overall Stereotypicality” was the only factor with loadings completely aligned with the originally proposed theory and solely contained items which assessed the perceived overall stereotypicality of an individual. These findings suggest that overall stereotypicality addresses something distinctly different from other items that assess individual features. These results also indicate that perceived overall stereotypicality is a meaningful dimension that demonstrates reliability, mirrors originally proposed theory and requires the least amount of revision.

It is important to note that respondents struggled to identify pictures that approximated their own features and did not choose images which strongly aligned with their self-perceptions of said features. This discrepancy led to no pictorial items loading strongly onto any factors. Surprisingly, though skin tone was qualitatively rated as the most prominent indicator of stereotypicality, skin tone related items did not strongly factor with other physical features of stereotypicality. Instead, all other physical features factored together (i.e., nose, lips, hair color, eye color, hair texture), which may suggest that skin tone should be assessed independent of other physical features, when attempting to measure Black stereotypicality.

## Limitations

A major limitation of this study was the use of Amazon Mechanical Turk, which led to mass amounts of unusable data, spam, and computer-generated data and may have negatively impacted the integrity of the data and the way that items loaded onto factors. Another limitation

was a lack of standardized wording when referring to respondent's features and when developing Likert type labels for items. Additionally, the gradations in size between pictorial images were not standardized, which may have contributed to participants' confusion when attempting to identify items that closely resembled their features. Finally, several features held by Black respondents are susceptible to change over time and may have led to confusion or inconsistent answers among respondents. For example, older Black individuals may perceive their "natural" hair color as Black, however, aging has whitened or peppered parts of their hair. Similarly, respondents with permanent hair styles (e.g., dreadlocks) may have permanently altered "natural" hair and be unsure of how to describe their current texture.

## Future Studies

Because skin tone and physical features related to Black stereotypicality have been associated with important outcomes ranging from low self-esteem and poor health outcomes to experiences of discrimination, more work should be done to understand how these constructs may function together. The results of this study provide a steppingstone which invites future researchers to revise and validate items pertaining to stereotypicality to better understand how and why various features cluster together. Future work to fully develop a functioning, weighted, measure will allow researchers to further investigate how various features influence the perceptions and experiences of Black individuals.

To accomplish this, a few guidelines should be noted based off this preliminary study. (1) Reverse coded items with polar opposite wording should be avoided during the preliminary stages of developing this scale, to avoid method-based factors from being formed and identified. (2) Since skin-tone items did not significantly load onto any factors, future researchers should consider viewing skin-tone as a separate construct and potentially utilize an established measure

of skin-tone to assess construct validity and better comprehend how these items relate. (3) All items should utilize a singular, uniform Likert label for each item (e.g., *strongly agree* to *strongly disagree*) to prevent method factors. (4) Future studies should also phrase questions to focus on either self-perceptions or the perceptions of other Black people to avoid confusion and prevent further methodological driven factors. (5) If pictorial items are used, an application like photoshop should be utilized to ensure that gradations between each image increase in a continuous and standardized fashion (E.g., the difference in width between nose 1 and nose 2 should be the same as the difference in width between nose 3 and nose 4). Platforms like Amazon's Mechanical Turk should be avoided to thwart data farmers and bots who may willingly deceive researchers to qualify for paid studies to the detriment of data quality. (6) "Overall stereotypicality" should be further tested and considered in future research as it is reliable, factors together, and ultimately homes in on perceptions that may have major implications for better addressing racial disparities and bias that occur in the real world (i.e., racial profiling, perceived discrimination, disparate health treatment) by highlighting how one's composite physical features may influence their experiences in a way that skin tone not cannot solely account for.

# APPENDIX

# COGNITIVE PRETESTING INTERVIEW PROTOCOL

**Perceived Black Stereotypicality Scale: Cognitive Interview Protocol**

**Interviewer**: Hello and thank you for agreeing to participate in this study today. The objective of this study is to create a scale which assesses an individuals' perceived "Black stereotypicality". This word refers to the extent to which a Black person physically looks like a member of their own racial group. For instance, an individual's skin-tone is an example of one feature that people use when trying to determine the racial background of others. This scale aims to assess physical, facial features beyond skin color that make a person look more or less Black. Does that make sense?

This interview will essentially occur in two parts. During the first section I will ask you to read aloud and answer questions from the actual survey. As you answer each question, I may ask additional follow up questions for clarity.

Once you have completed the survey, we will proceed to the second portion of the interview. During this part, I will ask you a series of questions about Black facial features which aid in identifying racial group membership for Black individuals. Do you have a basic understanding of our interview process for today?
So first, I'll send over the link to the PBS Scale. Afterward I will ask that you share your screen
Link: https://unt.az1.qualtrics.com/jfe/form/SV_2h2wsKHHNxXbZSm

So we will now start part I of our interview. Please read aloud and answer each question on the survey. If you come across a question that his hard to understand or confusing, please let me know.

Part I: Cognitive Survey Probes:

1. Did you find it easy or hard to use the pictures for identifying your own facial features?
2. **Does this question seem like a good indicator of 'Black Stereotypicality"?
3. **Was this question confusing?
    a. Would you re-word it in any other way?
4. How do you interpret the term "stereotypically Black"?

Part II: Structured Qualitative Interview:

5. What facial features are usually used to help identify Black racial group membership?
6. Take the time to describe to me what a "stereotypically Black" face looks like in your head
7. Please take a moment to imagine the face of a "typical" Black person. What distinct facial features does this face have that a "typical" White face may not have?
8. Please take a moment to imagine the face of a "typical" Black, multiracial person. What features make this person look more racially ambiguous?

| Dimensions | Conceptualization |
|---|---|
| I. Physiognomic Features (Nose and Lips) | Refers to the size of features on a person's face that have been historically associated with racial typicality. Specifically, it focuses on physical features, such as the broadness of an individual's nose and the fullness of their lips. |
| II. Hair Texture | Refers to how kinky/curly an individual's natural hair is |
| III. Feature colors/tones | Focuses on the colors and hues of an individual's physical features. Specifically, hair color and eye color are encompassed by this dimension. |
| IV. Skin Tone | Refers to the degree of lightness or darkness of an individuals' skin |
| V. Overall Perceived Stereotypicality | Refers to how individuals' degree of stereotypicality is perceived by themselves and others in their surrounding environment. |

9. On the screen you will see a chart which breaks up facial features that help determine Black racial group membership into five major categories. Please take a moment to look over each category.
    a. Are there any categories that you would change or delete entirely?
    b. What categories seem most relevant when determining one's Black racial group membership?
    c. What categories seem least relevant?
    d. If you were to rank the categories by relevance (from 1 to 5) how would you rank them?

10. Out of these two questions, which is worded in a way that is easier to understand and answer?

    On a scale of 1 (completely disagree) to 7 (completely agree), how much do you agree with the following statement: "As a Black individual, I believe that I look like a typical member of my racial group."

    OR

    On a scale of 1 (completely disagree) to 7 (completely agree), how much do you agree with the following statement: "As a Black individual, other Black people think I physically look like a member of my racial group"

REFERENCES


Allport G. W. (1954). *The nature of prejudice.* Perseus Books.

Adams, E. A., Kurtz-Costes, B. E., & Hoffman, A. J. (2016). Skin tone bias among African Americans: Antecedents and consequences across the life span. *Developmental Review, 40*, 93–116. https://doi.org/10.1016/j.dr.2016.03.002

Alon, L., Smith, A., Liao, C., & Schneider, J. (2019). Colorism demonstrates dampened effects among young Black men who have sex with men in Chicago. *Journal of the National Medical Association, 111*(4), 413–417. https://doi.org/10.1016/j.jnma.2019.01.011

Averhart, C. J., & Bigler, R. S. (1997). Shades of meaning: Skin tone, racial attitudes, and constructive memory in African American children. *Journal of Experimental Child Psychology, 67*(3), 363–388. https://doi.org/10.1006/jecp.1997.2413

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric features in criminal sentencing. *Psychological Science, 15*, 674–679. https://doi.org/10.1111/j.0956-7976.2004.00739.x

Blake, J.J., Keith, V.M., Luo, W., Le, H. & Salter, P. (2017). The role of colorism in explaining African American females' suspension risk. *School Psychology Quarterly, 32*(1), 118–130. https://doi.org/10.1037/spq0000173

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*(149). https://doi-org.libproxy.library.unt.edu/10.3389/fpubh.2018.00149

Branigan, A. R., Wildeman, C., Freese, J., Kiefe, C. I. (2017). Complicating colorism: Race, skin color, and the likelihood of arrest. *Socius: Sociological Research for a Dynamic World, 3*, 1-17. https://doi.org/10.1177/2378023117725611

Breland, A. (1998). A model for differential perceptions of competence based on skin tone among African Americans. *Journal of Multicultural Counseling and Development, 26*(4), 294–312. https://doi.org/10.1002/j.2161-1912.1998.tb00206.x

Breland-Noble, A. M. (2013). The impact of skin color on mental and behavioral health in African American and Latina adolescent girls: A review of the literature. In R. E. Hall (Ed.), *The melanin millennium* (pp. 219–229). Springer Science. https://doi.org/10.1007/978-94-007-4608-4_14

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues, 55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brown, K. T., Ward, G. K., Lightbourn, T., Jackson, J. S. (1999). Skin tone and racial identity among African Americans: A theoretical and research framework. In Jones, R. L. (Ed.),

*Advances in African American psychology: Theory, paradigms, and research* (pp. 191–214). Cobb Publishers.

Caldwell, K. L. (2003). “Look at her hair”: The body politics of Black womanhood in Brazil. *Transforming Anthropology, 11*, 18–29. https://doi.org/10.1525/tran.2003.11.2.18.

Clark, K. B. & Clark, M. B. (1947). Racial identification and preference in negro children. In E. L. Hartley (Ed.) *Readings in Social Psychology*. New York: Holt, Rinehart, and Winston.

Coard, S. I., Breland, A. M., & Raskin, P. (2001). Perceptions of and preferences for skin color, Black racial identity, and self-esteem among African Americans. *Journal of Applied Social Psychology, 21* (11), 2256-2274.

Cowart, K. O., & Lehnert, K. D. (2018). Empirical evidence of the effect of colorism on customer evaluations. *Psychology & Marketing*, *35*(5), 357–367. https://doi.org/10.1002/mar.21091

Dhillon-Jamerson, K. K. (2018). Euro-Americans favoring people of color: Covert racism and economies of White colorism. *American Behavioral Scientist*, *62*(14), 2087–2100. https://doi.org/10.1177/0002764218810754

Dixon, A. R. (2019). Colorism and classism confounded: Perceptions of discrimination in Latin America. *Social Science Research*, *79*, 32–55. https://doi.org/10.1016/j.ssresearch.2018.12.019

Dixon, T. L., & Maddox, K. B. (2005). Skin tone, crime news, and social reality judgments: Priming the stereotype of the dark and dangerous black criminal. *Journal of Applied Social Psychology, 35*(8), 1555–1570. https://doi-org.libproxy.library.unt.edu/10.1111/j.1559-1816.2005.tb02184.x

Dixon, A. R., & Telles, E. E. (2017). Skin color and colorism: Global research, concepts, and measurement. *Annual Review of Sociology*, *43*(1), 405–424. https://doi-org.libproxy.library.unt.edu/10.1146/annurev-soc-060116-053315

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing Outcomes. *Psychological Science, 17*(5), 383–386. https://doi-org.libproxy.library.unt.edu/10.1111/j.1467-9280.2006.01716.x

Foy, S. L., & Ray, R. (2019). Skin in the game: Colorism and the subtle operation of stereotypes in men’s college basketball. *American Journal of Sociology*, *125*(3), 730–785. https://doi-org.libproxy.library.unt.edu/10.1086/707243

Hagiwara, N., Kashy, D. A., Cesario, J. (2012). The independent effects of skin tone and facial features on Whites’ affective reactions to Blacks. *Journal of Experimental Social Psychology, 48*, 892–898. https://doi.org/10.1016/j.jesp.2012.02.001

Hairston, T. R., Laux, J. M., O'Hara, C., Roseman, C. P., & Gore, S. (2018). Counselor education students' perceptions of wellness and mental health in African American men: The effects of colorism. *Journal of Multicultural Counseling and Development, 46*(3), 171–185. https://doi.org/10.1002/jmcd.12100

Hamm, N. H., Williams, D. O., & Dalhouse, A. D. (1973). Preferences for Black skin among Negro adults. *Psychological Reports, 3*(2), 1171-1175.

Hannon, L., DeFina, R., & Bruch, S. (2013). The relationship between skin tone and school suspension for African Americans. *Race and Social Problems, 5*, 281–295. https://doi.org/10.1007/s12552-013-9104-z

Hannon, L. & DeFina, R. (2016). Reliability concerns in measuring respondent skin tone by interviewer observation. *Public Opinion Quarterly, 80*(2), 534-541. https://doi.org/10.1093/poq/nfw015

Harris, K. L. (2018). Biracial American colorism: Passing for White. *American Behavioral Scientist, 62*(14), 2072–2086. https://doi.org/10.1177/0002764218810747

Harrison, M. S., & Thomas, K. M. (2009). The hidden prejudice in selection: A research investigation on skin color bias. *Journal of Applied Social Psychology, 39*(1), 134–168. https://doi.org/10.1111/j.1559-1816.2008.00433.x

Harvey, R. D., Tennial, R. E., & Hudson Banks, K. (2017). The development and validation of a colorism scale. *Journal of Black Psychology, 43*(7), 740–764. https://doi.org/10.1177/0095798417690054

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238–247. https://doi.org/10.1037/1040-3590.7.3.238

Hebl, M. R., Williams, M. J., Sundermann, J. M., Kell, H. J., & Davies, P. G. (2012). Selectively friending: Racial stereotypicality and social rejection. *Journal of Experimental Social Psychology, 48*(6), 1329–1335. https://doi.org/10.1016/j.jesp.2012.05.019

Hersch, J. (2018). Colorism against legal immigrants to the United States. *American Behavioral Scientist, 62*(14), 2117–2132. https://doi.org/10.1177/0002764218810758

Hill, M. E. (2002). Skin color and the perception of attractiveness among African Americans: Does gender make a difference? *Social Psychology Quarterly, 65*(1), 77. https://doi.org/10.2307/3090169

Johnson, T. & Bankhead, T. (2014). Hair it is: Examining the experiences of Black women with natural hair. *Open Journal of Social Sciences, 2*(1), 86-100. https://doi.org/10.4236/jss.2014.21010

Kahn, K. B., & Davies, P. G. (2010). Differentially dangerous? Phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes & Intergroup Relations, 14*(4), 569–580. https://doi.org/10.1177/1368430210374609

Kahn, K. B., Goff, P. A., Lee, J. K., & Motamed, D. (2016). Protecting whiteness: White phenotypic racial stereotypicality reduces police use of force. *Social Psychological and Personality Science, 7*(5), 403–411. https://doi.org/10.1177/1948550616633505

Kahn, K. B., Lee, J. K., Renauer, B., Henning, K. R., & Stewart, G. (2017). The effects of perceived phenotypic racial stereotypicality and social identity threat on racial minorities' attitudes about police. *The Journal of Social Psychology, 157*(4), 416–428. https://doi.org/10.1080/00224545.2016.1215967

Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review, 37*(2), 214–233. https://doi-org.libproxy.library.unt.edu/10.1177/0894439317752406

Kleider-Offutt, H. M., Bond, A. D., Williams, S. E., & Bohil, C. J. (2018). When a face type is perceived as threatening: Using general recognition theory to understand biased categorization of Afrocentric faces. *Memory & Cognition, 46*(5), 716–728. https://doi.org/10.3758/s13421-018-0801-0

Kwate, N. O. A. & Threadcraft, S. (2015). Perceiving the Black female body: Race and gender in police constructions of body weight. *Race and Social Problems, 7*(3), 213-226. https://doi.org/10.1007/s12552-015-9152-7

Landor, A. M., Simons, L. G., Granberg, E. M., & Melby, J. N. (2019). Colorizing self-esteem among African American young women: Linking skin tone, parental support, and sexual health. *Journal of Child and Family Studies, 28*(7), 1886–1898. https://doi.org/10.1007/s10826-019-01414-8

Landor, A. M., & Smith, S. M. (2019). Skin-tone trauma: Historical and contemporary influences on the health and interpersonal outcomes of African Americans. *Perspectives on Psychological Science, 14*(5), 797–815. https://doi.org/10.1177/1745691619851781

Lee, M. R., & Thai, C. J. (2015). Asian American phenotypicality and experiences of psychological distress: More than meets the eyes. *Asian American Journal of Psychology, 6*, 242–251. https://doi.org/10.1037/aap0000015

Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of Black Americans: reexploring the role of skin tone. *Personality and Social Psychology Bulletin, 28*(2), 250–259. https://doi.org/10.1177/0146167202282010

Maddox, K. B., & Perry, J. M. (2018). Racial appearance bias: Improving evidence-based policies to address racial disparities. *Policy Insights from the Behavioral and Brain Sciences, 5*(1), 57–65. https://doi.org/10.1177/2372732217747086

Massey, D. S. & Martin, J. A. (2003). *The NIS Skin Color Scale*. Princeton University Press.

Opie, T. R., & Phillips, K. W. (2015). Hair penalties: The negative influence of Afrocentric hair on ratings of Black women's dominance and professionalism. *Frontiers in Psychology, 6.* https://doi.org/10.3389/fpsyg.2015.01311

Richeson, J. A., Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual Review of Psychology, 67,* 439-63. https://doi-org.libproxy.library.unt.edu/10.1146/annurev-psych-010213-115115

Russell, K. Y., Wilson, M., & Hall, R. E. (2013). *The color complex (revised): The politics of skin color in a new millennium*. Random House.

Ryabov, I. (2013). Colorism and school-to-work and school-to-college transitions of African American adolescents. *Race and Social Problems, 5*(1), 15–27. https://doi.org/10.1007/s12552-012-9081-7

Ryabov, I., & Goza, F. W. (2014). Phenotyping and adolescence-to-adulthood transitions among Latinos. *Race and Social Problems, 6*(4), 342–355. https://doi.org/10.1007/s12552-014-9132-3

Schinka J. A., Velicer W. F., Weiner I. R. (2012). *Handbook of psychology, vol. 2, research methods in psychology*. John Wiley & Sons, Inc.

Sims, C., & Hirudayaraj, M. (2016). The impact of colorism on the career aspirations and career opportunities of women in India. *Advances in Developing Human Resources, 18*(1), 38–53. https://doi.org/10.1177/1523422315616339

Stepanova, E. V., & Strube, M. J. (2009). Making of a face: Role of facial physiognomy, skin tone, and color presentation mode in evaluations of racial typicality. *Journal of Social Psychology, 14*9(1), 66–81. https://doi.org/10.3200/SOCP.149.1.66-81

Sweet, E., McDade, T.W., Kiefe, C. I., & Liu, K. (2007). Relationships between skin color, income, and blood pressure among African Americans in the CARDIA study. *American Journal of Public Health, 97*(12), 2253–2259. https://doi.org/10.2105/AJPH.2006.088799

Thompson, M., & Keith, V. (2001). The blacker the berry: Gender, skin tone, self-esteem, and self-efficacy. *Gender and Society, 15*, 336–357. https://doi.org/10.1177/089124301015003002

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219–246. https://doi-org.libproxy.library.unt.edu/10.1177/0095798418771807

Wilkins, C. L., Chan, J. F., & Kaiser, C. R. (2011). Racial stereotypes and interracial attraction: Phenotypic prototypicality and perceived attractiveness of Asians. *Cultural Diversity and Ethnic Minority Psychology, 17(*4), 427–431. https://doi-org.libproxy.library.unt.edu/10.1037/a0024733

Williams, M. J., George-Jones, J., & Hebl, M. (2019). The face of STEM: Racial phenotypic stereotypicality predicts STEM persistence by—and ability attributions about—students of color. *Journal of Personality and Social Psychology*, *116*(3), 416–443. https://doi.org/10.1037/pspi0000153

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology, 113*(1), 59–80. https://doi.org/10.1037/pspi0000092

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

Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PloS one*, *11*(6). https://doi.org/10.1371/journal.pone.0157795