Best Practices for Researching Diverse Groups

A. Kathleen Burlew, Bridgette J. Peteet, Caravella McCuistian, and Brittany D. Miller-Roenigk
University of Cincinnati

The elimination of health and other disparities requires high quality and methodologically sound research on racial/ethnic minorities. Despite a general consensus on the need for valid research on racial/ethnic minorities, few guidelines are available. This article contributes to closing this gap by discussing examples and strategies for addressing concrete issues that researchers may face during these stages of the scientific process: planning and literature review (identifying meaningful gaps and appropriate theoretical perspectives), design (caveats of race-comparison, selection of appropriate terminology), measurement (measurement equivalence, effects of ethnicity of the interviewer/coder), recruitment (barriers and strategies to facilitate recruitment), data analysis (use of norms derived from other groups, hazards of combining ethnic groups in the analyses), and dissemination of study findings to professional and lay audiences. Applying appropriate methodology will result in research that may impact disparities.

Public Policy Relevance Statement
The growing diversity of the nation increases the urgency for research that addresses health disparities. This article provides guidelines and recommendations for addressing issues encountered at each stage of the research project.

Effective research has the ability to potentially contribute to the development of a world without racial/ethnic health and other disparities. However, conducting research that minimizes health disparities is more complicated than simply adding more racial/ethnic minorities. Instead, as the National Institutes of Health (NIH, 2001), the American Psychological Association (APA, 2017), and other professional organizations have asserted, the way forward requires a critique of the usual scientific methods (e.g., measurement, design) to identify the most appropriate strategies for research on racial/ethnic minorities. Research conducted without cultural consideration can potentially result in incomplete or even incorrect assumptions that do little to reduce health disparities. However, high quality, methodologically sound research on racial/ethnic minorities may generate information that promotes health equity for underserved populations throughout the nation.

The growing diversity of the nation increases the urgency for research that addresses disparities. Existing research provides a compelling argument that findings based on predominately White samples may be insufficient for understanding other racial/ethnic minority groups. For example, Brody and colleagues (2004) demonstrated that including variables relevant for a specific racial/ethnic minority group added value to the research. Specifically, they reported that racial socialization, although not typically included in mainstream research on Whites, is important to consider in research on Black youth. To reduce health disparities, behavioral research may be required that includes concepts important to racial/ethnic minority groups (e.g., acculturation, racial identity) and/or examines, rather than assumes, the extent to which findings from one group generalize to a different group.

NIH has acknowledged the contribution of research to achieving health equity. The NIH (2016) Strategic Plan, 2016–2020 calls for the "Evidence-Based Reduction of Health Disparities." The NIH Guidelines on the Inclusion of Women and Minorities, although restricted to clinical trials, is consistent with a growing body of research suggesting that ignoring race/ethnicity is a barrier to reducing health disparities. Recent revisions to that policy stipulate that, despite the importance of enrolling members of racial/ethnic minority groups as research participants, inclusion alone is insufficient for addressing health disparities. Instead, the policy was amended in 2017 to emphasize and clarify the valid analysis requirement (NIH, 2017). However, even that requirement only stipulated separate analysis "by sex/gender and race/ethnicity." Future research aimed at understanding racial/ethnic minorities will require techniques beyond simply conducting separate analyses for each race/ethnicity.

A. Kathleen Burlew, Bridgette J. Peteet, Caravella McCuistian, and Brittany D. Miller-Roenigk, Department of Psychology, University of Cincinnati.

Correspondence concerning this article should be addressed to A. Kathleen Burlew, Department of Psychology, Mail Location 210376, University of Cincinnati, Cincinnati, OH 45221. E-mail: kathy.burlew@uc.edu

Only limited attention has been devoted to establishing guidelines for valid analyses when conducting research on racial/ethnic minorities. Even the standards on quantitative research published in a recent issue of the *American Psychologist* only mention the need to report ethnicity along with other sociodemographic characteristics of the sample (Appelbaum et al., 2018). Fortunately, the APA (2017) and various scholars (Burlew, Feaster, Brecht, & Hubbard, 2009; Burlew, Weekes et al., 2011; Helms, Jernigan, & Mascher, 2005; Sue & Dhindsa, 2006) have added to our understanding of the requirements for valid analysis. The goal of this article is to address this gap by discussing barriers to methodologically sound health disparities research encountered at each stage of a research project and to offer recommendations for addressing those challenges. The sections, consistent with the scientific method, include planning and literature review, design, measurement, recruitment, data analysis, and dissemination.

Step 1: Planning and Literature Review

The review of relevant literature and the identification of useful theoretical perspectives are essential prerequisites to designing a research project that contributes meaningfully to the existing body of work. We discuss both in this section.

Step 1A: Review the Relevant Research to Identify Existing Gaps

One aim of the literature review is to determine gaps worthy of further investigation. Among the multiple paradigms that Joireman and Van Lange (2015) propose for identifying innovative research or gaps, perhaps "exploring the role of culture" is the most relevant paradigm for this discussion. Culture might be defined as the shared traditions, customs, beliefs, history, values, norms, and behaviors that provide a group with a framework of living. Beyond race, which groups people solely according to similar biogenetic characteristics, culture more significantly impacts an individual's actions and beliefs (Fernando, 2010). Studies rooted in the cultural paradigm may examine whether psychological/behavioral/socio-ecological principles established within one group are applicable to a different racial/ethnic minority group or whether alternative principles are more appropriate. For example, although a meta-analysis by Yoon and colleagues (2013) provided convincing evidence that acculturation, a cultural variable, is a useful concept for understanding mental health among Asian Americans, acculturation may not be a useful concept for studying all groups.

Because the research on specific racial/ethnic minorities groups is typically limited, finding relevant literature on racial/ethnic minorities may be more of a challenge. We have several recommendations for researchers studying racial/ethnic minority groups. First, instead of limiting the literature review to their own fields, we encourage researchers to expand their literature review beyond their own disciplines. An interdisciplinary approach may be particularly useful for identifying relevant contextual factors that, if included, would increase the meaningfulness of the findings. Similarly, we caution researchers against restricting their research questions to examining whether the same relationships exist among variables for a specific racial/ethnic minority group that were previously demonstrated with White samples. Instead, we encourage researchers to explore research questions that consider the unique characteristics of their target group (e.g., immigration status, racial socialization). Finally, we also encourage researchers to consider strengths-based approaches rather than simply relying on deficit approaches. For example, Peteet and colleagues studied the characteristics of high-achieving Black youth (Peteet, Montgomery, & Weekes, 2015) rather than the predictors of academic failure commonly studied in the literature.

Step 1B: Identifying an Appropriate Theoretical Framework

Theories inform the research question, predict the expected direction of the relationship between variables, and assist in the interpretation of the results. Hall, Yip, and Zárate (2016) classify theoretical models in psychology into three orientations for studying culture: *generalizability, group-differences*, and the *multicultural perspective*.

A generalizability approach assumes that groups have more similarities than differences. Therefore, an underlying assumption is that a theory that does not generalize across groups may not be a valid theory. If the researcher is proposing to show that a universal principle applies to a specific target group, then selecting a theory with some evidence of demonstrated applicability across groups (e.g., Social Learning Theory) may be most appropriate. Alternatively, the group difference approach may be most appropriate if the researcher is interested in identifying cultural mechanisms that may explain why two groups differ on a construct. Finally, if the researcher is examining the role of a variable unique to a specific target group (e.g., racial identity), choosing a theory consistent with a multicultural approach may be best.

The three theoretical approaches (generalizability, group differences, and multicultural) all have advantages and disadvantages. The advantage of the generalizability approach is an identification of commonalities of the human experience; however, the inherent assumption of the homogeneity of the human experience and the disregard of group difference is a limitation of the generalizability approach (Hall et al., 2016). The strength of the group differences approach is its focus on exploring whether cultural mechanisms account for group differences; the disadvantage is that the focus may not include the unique contextual factors of specific groups and the approach may support the ethnocentric focus on the White comparison group as the standard (Hall et al., 2016). One strength of the multicultural approach is the focus on just one target group; however, the potential disadvantage is that ethnicity-focused models may promote stereotypes (Hall et al., 2016).

We recommend that researchers consider the stage of the research when deciding on the most appropriate research approach. Testing a generalizability theory might be a good start to see if a principle assumed to be universal applies to the target group. However, later in the research endeavor, the group differences and multicultural approaches may be more appropriate. For example, in a research program on sexual aggression among Asian American males, Hall and colleagues (Nagayama Hall, Teten, DeGarmo, Sue, & Stephens, 2005; Hall, DeGarmo, Eap, Teten, & Sue, 2006) first investigated whether a universal multifactorial model explained sexual aggression in both Asian Americans and Whites (generalizability). However, their research program later identified a cultural mechanism, loss-of-face (i.e., experiencing social hu-

miliation), as a deterrent to self-reported sexual aggression among Asian American, but not White men (group differences). Still later, a multicultural research project focused on within-group variability in sexual aggression (Hall et al., 2006; Nagayama Hall et al., 2005).

Step 2: Design

Once the research question(s) are set, the next step is to design a study appropriate for addressing those research questions. This section describes two issues facing research on racial/ethnic minorities in the design phase—deciding if a race-comparison design is appropriate and use of appropriate terminology.

Step 2A: Be Aware of the Caveats of Race-Comparison Designs

A typical race-comparison study assesses similarities and differences between two or more groups on a specific variable or outcome. Graham (1992), in a frequently cited classic study of six leading psychology journals, reported that 72% of the publications that studied Blacks were race-comparison studies. Even when the author is investigating an issue especially relevant for a specific racial/ethnic minority group, many journal editors still ask whether a comparison group is needed to clarify whether the findings differ from Whites. For example, one of the coauthors of this article once had a reviewer from a mainstream journal ask about a White comparison group when she submitted an article on sickle cell disease among Blacks. This question was particularly unexpected since the U.S. sickle cell population is mainly Black. The race-comparison designs that we are discussing here differ in one important way from the group difference approach discussed earlier. Whereas the emphasis of the group difference approach is on cultural mechanisms that explain group differences, race-comparison designs typically lead to studies that focus solely on the difference without exploring the role of cultural mechanisms. Although race-comparison designs may be appropriate in certain situations, we raise several concerns for researchers considering a race-comparison study.

A first concern is that race-comparison designs may, even if inadvertent, disregard meaningful within-group differences. A study by Weaver, Himle, Taylor, Matusko, and Abelson (2015) illustrates the importance of considering within-group differences. Their study of women and depression found that rural Black women had significantly lower odds of meeting criteria for a Major Depressive Disorder than urban Black women. Consequently, any race-comparison study of Black and White women on depression that groups all Black women together would likely ignore this difference. Clearly, increased immigration rates often increase within group heterogeneity (e.g., African Americans vs. Black Caribbeans) that may be overlooked in race-comparison designs.

The second concern, measurement equivalence, is discussed in more detail later in this article. However, the Child Behavior Checklist (CBCL) offers a good example of items that may lead to false conclusions because the items may not be as relevant for one group as another. Specifically, the assessment of social competence solicits information on participation in formal organizations (e.g., sports teams). However, a child in a family with limited financial resources may not be able to express his or her social competence in ways tapped by the CBCL. In that case, group differences on such items may be attributable to differences in opportunities rather than to actual differences in social competence. When the sociodemographics of the standardization sample differ markedly from the new target group, we recommend piloting the instrument ahead of time on the target group or conducting random probes (see discussion in Measurement section on qualitative methods) to identify statements that differ in their meaning across groups.

A third concern is that race/ethnicity may be a proxy for other differences. Sociodemographic factors such as income, education, socioeconomic status, background experiences, and access to opportunities may be so closely associated with race/ethnicity that race may operate as a proxy for sociodemographic factors (i.e., observed racial/ethnic differences may be attributable to other factors). A study by Carvalho et al. (2015) illustrates this race-as-a-proxy concern. They demonstrated that what might appear to be inherent racial differences in the cognitive decline of older adults may actually be largely attributable to literacy rates stemming back to the quality of early-life education (e.g., the resources that the school district spent on education, the length of the school year, and the teacher–student ratio) when that generation was attending school.

In addition, race/ethnicity can also be a proxy for contextual factors such as site differences. For example, if one racial/ethnic minority group in a treatment study were recruited from a public (and poorly) funded clinic but the other group were recruited from a private (and better) funded clinic, what may appear to be racial differences may be due to differences in site characteristics (e.g., differences in the quality of care).

The fourth concern is that race-comparison designs sometimes have a way of changing the research question or focus. Perhaps, inadvertently, the White sample becomes the gold standard against which the other group is evaluated. Imagine a hypothetical study originally intended to evaluate the efficacy of a specific treatment for a specific racial/ethnic minority group. However, the research team is convinced that the work would be viewed more favorably for publication if a White sample were included. Too often, once a White comparison group is added, the focus then switches to racial/ethnic minority differences in outcomes rather than the original question of the treatment's effectiveness for a specific group. Conceivably, within-group designs such as moderator variable, mediator variable, mediated-moderator (a third variable mediating the interaction effects of two other variables), and moderated-mediator (group differences in the mediating processes) studies may yield more useful information about the specific racial/ethnic minority group than race-comparison studies.

Response style differences across racial/ethnic minority groups may be a fifth source of concern in race-comparison research. For example, in a classic study by Bachman and O'Malley (1984), Black youth were more likely than White youth to use the extreme response options (*strongly agree* or *strongly disagree*) instead of more moderate options (*agree* or *disagree*) on Likert-type items. Similarly, within some Latino groups, *yea-saying* (i.e., saying yes regardless of the statement) is so common that a term (si-ismo) actually exists for it (Guerra & Jagers, 1998). Clearly, such cultural differences in response styles may be a confound in race-

comparison designs by contributing to the presence of racial/ethnic minority group differences on self-report measures.

Despite the cautions about race-comparison studies, these designs may be appropriate in specific situations. In a classic essay, Azibo (1988) identified three situations in which race-comparison might be appropriate. These included (a) refuting prevailing but untrue negative beliefs about a specific racial/ethnic minority group (e.g., disproving that Group X uses more drugs than other groups); (b) detecting racial differences on social/health indices in order to impact social policies (e.g., verifying disparate police stops; documenting housing or employment discrimination); or (c) comparing groups on worldview (e.g., Group X has more of an individualistic worldview but Group Y has more of a collectivistic worldview).

More recently, other scholars have pointed out additional situations where comparison may be appropriate. Joireman and Van Lange (2015) pointed out that comparison is fundamental to cross-cultural research aimed at exploring the role of culture in explaining similarities and differences in thought and behavior. Hall et al. (2016) contend that a group difference approach might be conceptualized as a prelude to race-specific research. However, they maintain that the next step after group differences are revealed is to determine the potential cultural mechanisms that account for the differences. Nevertheless, we offer an additional perspective. Specifically, even when similarities are revealed on findings, the underlying contextual factors associated with the finding may differ.

Overall, we contend that race-comparison designs may be appropriate when identifying similarities and differences pertinent to the research question. However, we recommend that researchers consider the concerns addressed above (i.e., appropriateness of a within-group design, measurement equivalence, whether race is a proxy, whether the design is consistent with the research question, and response style issues) when deciding whether a race-comparison study is appropriate.

Step 2B: Using Appropriate Terminology

Selecting suitable terminology in racial/ethnic scholarship is complex due to the historical, social, and political implications of labels (Bhopal & Donaldson, 1998) and the fact that many terms are based on cultural, phenotypic (e.g., skin tone, facial features), and genealogical characteristics rather than science (Cornell & Hartmann, 1998). Despite the importance of appropriate terminology to other diverse groups as well, we limit our discussions here to race/ethnicity. Research by Janet Helms and colleagues (2005) highlights the imprecision and hidden psychological constructs of racial categorization. In spite of these limitations, research participants are frequently forced to choose a label based on predetermined categories set by the government and replicated by researchers. Several admonitions for inquiring about race/ethnicity are described in this section.

Many researchers use the categories from the U.S. Census to define the study population. The 2010 Census included two questions with 14 options for race and five options for ethnicity including non-Hispanic and four Hispanic/country-of-origin options (U.S. Bureau of the Census, 2010). Although the options have broadened, this two-question approach contains some interesting assumptions. In particular, the approach differentiates among Asian (Japanese vs. Chinese), American Indian/Alaska Native (tribe), and Latino (Puerto Rican vs. Cuban) subgroups of different heritages but does not differentiate among Blacks who also have diverse nationalities. For example, Caribbean/West Indians (singer Rihanna), South Africans (former South African President, Nelson Mandela), and U.S.-born Blacks (former first lady Michelle Obama) would all mark the same response on the Census. Since the appropriate racial differentiations may vary based on factors such as the research question and the expected sample composition, we encourage researchers to pay attention during the design phase to determining the most appropriate approach to asking about race/ethnicity in their particular study.

Race/ethnicity is most commonly self-defined, but researchers sometimes use proxies such as parent reports, parental race, and categorization by inference (e.g., birth/death certificate, observational research; Sandefur, Campbell, & Eggerling-Boeck, 2004). The NRC (2004) outlines four best practices for capturing race/ethnicity data: (a) allow participants to self-select; (b) allow participants the option of selecting multiple race/ethnicities; (c) ask about the country of origin (e.g., Cubans for Latinos); and (d) oversample specific rather than broad subgroups (e.g., Korean vs. Asian American).

If no other information is available, investigators might use past research to identify the terms preferred by the group of interest. Taylor, Lopez, Martínez, and Velasco (2012) found that Hispanic individuals prefer to be identified by their country of origin (e.g., Cuban, Mexican) and secondarily slightly prefer Hispanic over Latino. A new movement toward gender inclusivity has led to the use of the term *Latinx*, which is gaining popularity. Asian Americans express similar preferences for using their country of origin to label them (Census Briefs, 2012). Americans of African descent are divided equally in terms of their preference for the terms African American and Black. However, their preferences vary based on demographic factors such as age (older adults prefer Black), geography (people from larger and nonsouthern cities prefer African American; Sigelman, Tuch, & Martin, 2005), and acculturation (first-generation Caribbean immigrants do not necessarily identify with either term, but prefer Black over African American; Waters, 1999). Indigenous populations prefer tribal affiliation although an older survey indicated that nearly half (49.76%) preferred American Indian (Bureau of Labor Statistics, 1995) and the indigenous people of Alaska generally favor Alaska Native (NRC, 2004). For multiracial individuals, now one in four Americans (Taylor et al., 2012), offering multiple selection options is best, although the U.S. Bureau of the Census (2001) found that over 97% will select one race if not given the option to select more than one.

In addition to using appropriate labels to describe racial/ethnic minority groups, researchers should strive to avoid bias and/or stereotypical language in their professional writing and in the lay dissemination of their work. Several specific points are worth considering. First, although the literature is replete with conflicting definitions of cultural terminology, researchers should avoid outdated or offensive terms (e.g., *colored; Indian; illegals*, etc.) and instead, use more appropriate terms (e.g., *person/people of color, Native American, undocumented*, etc.; Substance Abuse & Mental Health Services Administration [SAMHSA], 2014). Second, researchers should follow the publication style guidelines of the respective field (e.g., capitalize race/ethnicity labels; APA, 2009).

Third, terms that imply a subordinate status or have subtle negative undertones should be avoided. For example, the term *minority* although commonly used to describe people of color, is an inaccurate description of global diversity. Similarly, many consider the term *non-White* depreciatory because non-White is not a race, ethnicity, or culture, and suggests that Whites are the normed reference group. Just as females are not called *non-males*, labeling racial minority ethnic groups by what they are not seems inappropriate. *Mulatto* is an antiquated term used to describe individuals with mixed ancestry (generally one Black and one White parent), but the term was originally used to describe a mule or the offspring of a horse and a donkey (Fly, 2010) and the term also excludes other mixed-race combinations.

Researchers should also be careful of *dog whistle language* (subtle and negative). For example, the title alone of an article on perceived attractiveness that appeared in *Psychology Today* provoked controversy. Moreover, the title "Why are Black Women Less Physically Attractive than Other Women?" although it may have appeared scientific on this reputable platform, invited criticism of the scientific merit of the work (Kaufman & Wicherts, 2011). This coded language includes words that are so often used to describe a group of people that the original meaning becomes permanently associated with that group. For example, using the word *ghetto* to describe someone has an undercurrent of negative perceptions of working-class or lower-income Blacks. Conversely, some perceive that even seemingly positive descriptions of specific members of certain groups such as *articulate* and *different* subtly suggest that the person is exceptional because the group lacks overall those positive characteristics. In sum, identification and selection of appropriate terminology is an important step in culturally competent research. We encourage researchers to consider the preferences of the particular group when asking about race/ethnicity and to attend to the potential for unintentional bias when reporting on those groups.

Step 3: Measurement

When selecting appropriate measures, researchers might consider three factors including: (a) the relation of cultural norms and practices to the relevance of specific constructs for a particular group; (b) the range of behaviors and responses that are indicators of the construct and; (c) how individuals understand and interpret items intended to assess the constructs (Hughes & DuMont, 1993). Accordingly, the adequacy of a measure for one culture or subculture does not guarantee the appropriateness of that measure for another cultural group (Bravo, 2003). Section 3A discusses issues essential to good measurement in racial/ethnic minority research using both quantitative and qualitative strategies. Section 3B also describes the effects of the ethnicity of both the interviewer and the coder on measurement.

Step 3A: Establish Measurement Equivalence

Although multiple definitions of measurement equivalence are available in the literature, a description suggested by Allen and Walsh (2000) is very appropriate for our purpose. Specifically, those researchers proposed that measurement equivalence exists when a scale measures "the equivalent underlying psychological construct in a new group or culture as the test measured within the original group in which it was standardized" (p. 67). McHorney and Fleishman (2006) argue that "attention to measurement equivalence is not an esoteric, psychometric issue that has little or no consequences for science, policy, or medicine. [Instead] understanding and assessing measurement equivalence is fundamental to science—to developing outcomes instruments, to theory building, (and) to testing hypotheses . . ." (p. S205).

Because the samples for much of behavioral research are White (Burlew, Larios et al., 2011), researchers studying racial/ethnic minority groups are especially likely to encounter measurement equivalence as a challenge. Bravo (2003) proposes that, even if two racial/ethnic minority groups do not differ on an underlying trait, the two groups may differ on their overall responses to a measurement item for multiple reasons including the following: (a) the situations mentioned in a measure may apply to one group but not another; (b) various cultural groups may differ in the connection between specific behaviors and the underlying trait; (c) differences in the opportunity structure may lead to differences in the manner in which certain traits are manifested; and (d) group differences in circumstances may result in differences in the meaning of a specific statement or behavior.

Although methods associated with classical test theory (e.g., reliability, validity) are commonly reported to defend the appropriateness of a measure, more sophisticated quantitative techniques are available including confirmatory factor analysis (CFA), item response theory (IRT), and functional equivalence.

Confirmatory factor analysis (CFA). Using CFA, a researcher can conduct a series of tests to examine whether a scale developed for one racial/ethnic minority group is assessing a similar construct in a different cultural group. Beyond the test of the equality of covariance matrices, Vandenberg and Lance (2000) describe multiple tests for assessing measurement equivalence. *Configural equivalence*, a test to determine whether the factor structure is similar across groups, is typically assessed first because configural equivalence is a necessary condition for the other tests. In Case Study #1 (configural example), the tests for configural equivalence revealed that the original factor structure of the Trauma Symptom Checklist-40 was not a good fit for the low-income substance-abusing women (Ghee, Johnson, & Burlew, 2010).

Case Study #1

Configural Equivalence Example

Ghee and colleagues (2010) investigated whether the Trauma Symptom Checklist-40 (TSC-40), originally developed for college-educated and professional women, was appropriate for economically and educationally disadvantaged substance-abusing women. The sample included 50 Black and 52 White women enrolled in residential substance use treatment. The results revealed that the original factor structure did not fit the sample of substance-abusing women. Since configural equivalence is a prerequisite for the other tests of measurement equivalence, instead of conducting additional measurement equivalence tests, the research team conducted exploratory factor analyses to construct an alternative but more appropriate factor structure for the substance abuse sample.

Configural Equivalence Example (continued)

Additional analyses revealed that the alternative factor structure was appropriate for both Black and White substance abusing women, suggesting configural equivalence of the alternative model for Black and White substance-abusing women.

If configural equivalence is supported, *metric equivalence*, similarities in the factor loadings, is frequently the second test of measurement equivalence. In Case Study #2 (Metric Equivalence Example), Tuliao, Landoy, and McChargue (2016) were able to demonstrate that both the factor structure (configural equivalence) and the factor loadings (metric equivalence) of the Alcohol Use Disorder Identification Test (AUDIT) were similar for U.S. and Filipino college students.

Case Study #2

Metric and Scalar Equivalence Example

Previous research suggests that the factor structure of the Alcohol Use Disorders Identification Test (AUDIT), a measure of hazardous drinking and alcohol use practices, varies across populations. Accordingly, Tuliao and colleagues (2016) investigated the measurement equivalence of the AUDIT for a U.S. ($n = 1,259$) and a Philippine population ($n = 255$). The initial confirmatory factor analyses demonstrated that the existing three-factor model was the best fit for both the U.S. and Philippine participants. The researchers then assessed for metric equivalence by comparing the factor loadings for each item on the AUDIT. They found that the factor loadings were similar on all items except Item 7 (feeling guilt or remorse after drinking), suggesting general metric equivalence between the two groups. However, an additional model that further constrained the two groups on invariant item intercepts (scalar nonequivalence) did not prove to be a well-fitting model. That finding suggested that the mean item scores were non-invariant, or varied, between the two groups. Additional analyses revealed that the U.S. sample had higher mean item scores on certain items and the Philippines sample had higher mean item scores on other items.

According to Meredith and Teresi (2006), weak equivalence has been established if a measure demonstrates both configural and metric equivalence. According to that criterion, the AUDIT has demonstrated weak equivalence for U.S. and Filipino college students. However, in addition to meeting the criteria for weak equivalence, the criteria for strong equivalence and strict equivalence require the groups to demonstrate invariant item intercepts or scalar equivalence and invariant error variance, respectively (Meredith & Teresi, 2006). Although a total of eight CFA tests of invariance are available, it would be unusual for a researcher to conduct all eight tests. However, researchers may at least test for configural equivalence since configural nonequivalence is generally assumed to be the most serious source of measurement nonequivalence (Vandenberg & Lance, 2000).

Item response theory (IRT). A fundamental assumption underlying IRT is that if two groups are responding similarly to a scale, the relation of each item to the overall scale score should be similar across groups. Accordingly, IRT assesses whether individuals from multiple groups matched on an overall trait (i.e., scale score) respond similarly to specific items of the scale. *Differential item functioning* (DIF) is evident when the plotting of the relationship of the item to the overall trait (i.e., the Item Characteristic

Curve) suggests that an item is more discriminating or more extreme in one racial/ethnic minority group than in another. In Case Study #3 (IRT example), the fact that Whites and Latinos and also Whites and Blacks had differential item functioning was evidence of measurement nonequivalence.

Case Study #3

Item Response Theory Example

Wu and colleagues (2010) completed a study with 682 cocaine users to explore racial/ethnic differences in response to items (i.e., differential item functioning) on the DSM-IV checklist for cocaine. After confirming that all cocaine dependence items were measuring the same factor, differential item functioning was assessed. Latinos with similar scores as Whites on the DSM-IV checklist for cocaine were more likely to endorse the tolerance item than Whites but less likely than Whites to endorse the "continued use despite problems" item. Similarly, Black participants who had similar overall scores as Whites were more likely than the White participants to endorse the "inability to cut down" item. These differences argued against measurement equivalence between Latinos and Whites and between Blacks and Whites.

Fortunately, if the analyses suggest group differences, the researcher can conduct further analyses to detect the item(s) responsible for the racial/ethnic minority group differences. Depending on the research question, once the items are identified, the researcher can determine the appropriate steps for addressing the nonequivalence (e.g., transformations, removal of the nonequivalent item[s], etc.).

Although both CFA and IRT are acceptable approaches to assessing measurement equivalence, the researcher has the challenge of deciding which strategy is more appropriate for a specific study. The passage below by Meredith and Teresi (2006) may be helpful for making that decision: "Although factor analysis is often applied to continuous observed variables, IRT (a nonlinear model) has traditionally been applied to categorical (e.g., dichotomous) observed variables (p. S74)."

Functional equivalence. Even if the CFA and IRT findings suggest that two groups are responding similarly to a scale, the relation of a scale score to some measure(s) external to the scale may differ across groups. This is known as *functional nonequivalence*. In the construct validity study by Knight, Little, Losoya, and Mulvey (2004) described in Case Study #4 (functional equivalence example), differences in the relationship of the Self-Report of Offending scale (SRO) to external variables (e.g., arrests, gang memberships) between White and Latino males and between White and Black males led the researchers to caution against assuming the functional equivalence of the SRO for those two groups.

Case Study #4

Functional Equivalence Example

Knight and colleagues (2004) completed a study evaluating the measurement equivalence of the Self Report of Offending (SRO) measure for 1,338 juvenile offenders of Latino, African American, and White backgrounds. Additional information was gathered from the participants regarding other constructs related to offending including gang membership, level of impulsivity, moral

---

Functional Equivalence Example (continued)

---

disengagement, parental monitoring, peer delinquency and antisocial influence, and level of social support. First, the researchers confirmed the item equivalence of the SRO across both gender and race. Next, the functional equivalence of the SRO was examined in a series of construct validity analyses that examined the relation of SRO to a set of external constructs related to offending. Regarding gender, the predicted relationships between SRO and these constructs were similar for males and females, suggesting functional equivalence. The similar relationship between the SRO and the other constructs was also evidence of the functional equivalence of the SRO for White and Latino juvenile offenders. However, although the relationship between the SRO and nine of the variables was similar for Black and White males, the relationships were different on four of the constructs. Consequently, the evidence for the functional equivalence of the SRO was not as compelling for Whites and Blacks as for Whites and Latinos.

Qualitative methods.   Although the emphasis of this section is on quantitative methods of evaluating measurement equivalence, qualitative methods are also useful for evaluating the appropriateness of a measure for a specific group. Along with focus groups, Teresi, Stewart, Morales, and Stahl (2006) list other useful qualitative methods including random probes. (e.g., interviews with respondents regarding their interpretation of specific survey/questionnaire questions), cognitive interviewing (i.e., asking respondents to verbalize their thoughts as they respond to specific questions), and interaction analysis (i.e., a review of audiotaped or videotaped discussions conducted with research participants during the pretests to detect differences in the interpretation of items across ethnic or language groups).

Cross-cultural translation and interpretation.   The importance of using appropriate language also extends to measurement as well. As the country grows more diverse, the native language for many research participants will be something other than English. For example, a third of Latinos speak Spanish as their primary language and another third are bilingual (Taylor et al., 2012). Some researchers mistakenly believe that simply using software such as Google Translate to translate an instrument is sufficient. However, translation equivalence is only achieved when the meaning of the items is similar across groups (Brislin, 1993).

Back translation is a common approach to linguistic equivalence. In back translation, a bilingual person translates the document into the target language. Second, a different bilingual person translates the text back to the original language. Next, the team compares the two versions to determine if they are equivalent. However, Douglas and Craig (2007) point out the limitations of back translation. In particular, the method may not capture the subtle meaning of nuances or idioms. Instead, they argue for a committee-based approach that includes several independent translations, a review to identify discrepancies and create a final version, and even an independent person to adjudicate or harmonize the two versions to ensure the meaning is similar in both languages. In addition to translation, instruments may need to be renormed and validated to ensure that the underlying constructs are unchanged.

Language is not the only factor resulting in group differences in the interpretation of assessment or research measures. Kim and

Zabelina (2015) add that cultural values, beliefs, experiences, communication patterns, teaching and learning styles, and epistemologies may influence the interpretation of test (and questionnaire) items. For example, Kim, DeCoster, Bryant, and Ford (2016) asserted that the measurement nonequivalence on one specific scale of psychological distress, the K6, was attributable to group differences in the interpretation of words such as *nervous* or phrases such as "everything an effort."

We encourage researchers to assess, rather than assume measurement equivalence. Moreover, future researchers would benefit from supplemental publications on the measurement equivalence of specific measures. We also recommend the use state-of-science techniques when translation is required.

Step 3B: Considering the Effects of the Ethnicity of the Interviewer and the Coder on Measurement

A longstanding body of research has demonstrated that characteristics of the interviewer influence measurement (Davis, 1997; Davis & Silver, 2003; Johnson & Parsons, 1994). The classic Social Desirability Theory posits that respondents will mask their true feelings and behaviors to appear to conform to social norms, to provide an answer that is perceived as correct, or to conform to the perceived preferences of the interviewer (Maccoby & Maccoby, 1954). For example, Davis (1997) found that Blacks tend to conceal their real feelings with White interviewers and speak more openly with Black interviewers. Similarly, Johnson and Parsons (1994) also found that race-matched interviewer–respondent dyads display less inhibited communication. Although these findings are especially true for in-person interviews, the same pattern has been demonstrated in telephone interviews when respondents are asked to report the perceived race of the interviewer based on the interviewer's voice (Davis & Silver, 2003; Moorman, Newman, Millikan, Tse, & Sandler, 1999; Stokes & Yeh, 1988). Case Study #5 provides a somewhat extreme example demonstrating the effect of the perceived race of the interviewer even on performance (i.e., stereotype threat). Accordingly, we encourage researchers to examine the race of the interviewer effect when appropriate, possibly in a secondary study.

Case Study #5

Exploring the Race-of-Interviewer Effect

Davis and Silver (2003) explored how stereotype threat and the race of the interviewer influenced a participant's performance on a political knowledge questionnaire. Interviewers of various racial backgrounds presented participants with questions about political knowledge (i.e., how many years is the term of office for a senator?) over the telephone. Following the questions, the interviewer asked the participants "What do you think is my racial background?" The researchers found that Blacks answered fewer answers correctly when they perceived the interviewer was White than when they perceived the interviewer was Black. For White participants, the scores did not vary based on the race of the interviewer. This research, although aimed at illustrating stereotype threat, also demonstrates a race-of-interviewer effect.

Stereotype threat maintains that the pressure to disconfirm and to avoid being judged by negative and potentially degrading ste-

reotypes interferes with the processing of information. We argue that the survey context contains many parallels to a testing environment in which stereotype threat might alter responses to factual questions. Through a series of framing experiments in a public opinion survey and the reliance on the sensitivity to the race of the interviewer, our results are consistent with expectations based on a theory of stereotype threat. African American respondents to a battery of questions about political knowledge get fewer answers right when interviewed by a White interviewer than when interviewed by an African American interviewer. The Observed Differences in Performance on the Political Knowledge questions cannot be accounted for by differences in the educational background or gender of the respondents.

The race/ethnicity of the coder also may influence the data. For example, Gonzales, Cauce, and Mason (1996) instructed Black and White research assistants to code taped interactions of Black mother–daughter dyads. Although all coders were trained on the parenting styles of Black families, the White coders still reported more combative interactions and controlling behaviors than the Black coders. Conversely, the Black and White coders in a study by Melby, Hoyt, and Bryant (2003) rated families more favorably in the other race than families that matched their own race. Ironically, the authors noted an inverse relationship between rater bias and training. That is, training increased rather than decreased bias. Clearly, the evidence suggests that the race of both the interviewer and the coder should be considered rather than ignored in the design stage. Similar to measurement equivalence, we encourage researchers to advance the science by designing and publishing supplemental articles examining whether the race of either the interviewer or the coder impacted the findings.

Step 4: Recruitment

The next step is to utilize effective recruitment strategies to increase racial/ethnic minority participation in research. This section includes a discussion on barriers to recruitment faced by racial/ethnic minorities and a review of effective and ineffective recruitment strategies.

Step 4A: Barriers to Recruitment of Racial/Ethnic Minorities

Recruitment of racial/ethnic minorities can be challenging because these groups often face numerous barriers and limitations that reduce their participation in research and, consequently, result in their underrepresentation in study samples (Waheed, Hughes-Morley, Woodham, Allen, & Bower, 2015). Brown, Marshall, Bower, Woodham, and Waheed (2014) describe five types of barriers: (a) participant related, (b) practical issues, (c) family/community related, (d) health services related, and (e) research process. The resulting underrepresentation of racial/ethnic minority groups leads to fewer interventions with demonstrated effectiveness for racial/ethnic minority populations (Waheed et al., 2015). Compliance with the NIH Guidelines on the Inclusion of Women and Racial/Ethnic Minorities (NIH, 2001) will address this underrepresentation, aid in valid analyses, improve intervention outcomes, and contribute to the reduction of disparities for racial/ethnic minority groups. However, compliance requires the development of effective strategies to overcome barriers to the recruitment of racial/ethnic minority groups (Waheed et al., 2015). We recommend that researchers consult the existing literature summarized in the next section to identify effective recruitment strategies for their target group.

Step 4B: Identify Strategies to Facilitate Recruitment of Racial/Ethnic Minorities

Addressing these barriers requires the identification of effective methods and strategies (Burlew, Larios et al., 2011). In an effort to identify effective recruitment strategies of racial/ethnic minorities, we conducted a PsycINFO search using combinations of keywords: "recruitment AND racial/ethnic minorities." Although this search identified 98 articles, nearly three-quarters of the articles were eliminated because they either were not studying racial/ethnic minorities in the United States or they only identified barriers to recruitment, rather than reporting successful strategies to increase recruitment. Twenty-three articles ultimately were included in this review of the literature (see Table 1).

Effective versus ineffective methods of recruitment. The literature review outlines various methods of recruitment and their efficacy in recruiting racial/ethnic minorities, which are included in Table 1. Overall, the literature review was consistent with the conclusion in an earlier review by Yancey, Ortega, and Kumanyika (2006) suggesting that community involvement and outreach methods are more effective than other methods in alleviating barriers and increasing recruitment. This pattern was evident for Blacks, American Indian/Alaska Natives, and Latinos. As Table 1 details, the evidence supported the efficacy of several specific community recruitment methods for improving racial/ethnic participation including community outreach, Respondent-Driven Sampling, Community-Based Participatory Research, and recruiting at locations within the target group's community.

The literature review also suggested that methods that did not involve the community were less effective. As Table 1 details, the less effective methods included Internet/online social networks, media resources, adding health information about the target group to the recruitment materials, insertion of a picture of the racial/ethnic minority researcher to recruitment materials, including a culturally sensitive letter stressing the need for more information on the target group, use of referral sources, increasing recruitment sites, and using recruiters from the same race/ethnic group as the target. Although the evidence suggested these methods were generally ineffective, use of Internet/online social networks, advertising in media outlets, and using recruiters from the same race/ethnic group as the target were sometimes effective. However, overall, allocating resources to implement these methods may not be feasible. More research is needed to identify effective recruitment methods.

Differences between racial/ethnic minority groups. Some recruitment methods are effective for one racial/ethnic minority group, but not another. For example, Respondent-Driven Sampling is a method in which participants within a given study agree to recruit participants within their own social networks (e.g., family and friends; Burlew, Larios et al., 2011). Although

Table 1. *Community and Noncommunity Recruitment Strategies for Racial/Ethnic Minorities*

Community-based methods	Non-community-based methods
Community outreach + Black; weight loss (Hartlieb et al., 2015) + Black & Latino; HIV positive women recovering from Substance Use Disorder (Alvarez, Vasquez, Mayorga, Feaster, & Mitrani, 2006) + Black; smoking cessation (Okuyemi et al., 2007)	Internet/Online social networks + Latino; HIV/STI study (Fernández et al., 2004) − Latino; HIV positive prevention intervention (Hatfield et al., 2010)
Respondent-driven sampling + Black; injection drug using (Robinson et al., 2006) + Black; adolescents with ADHD and Substance Use Disorder (Holmes, Pressley, Haynes, Tyson, & Riggs, n.d.) + Black; HIV positive prevention intervention (Hatfield et al., 2010)	Use of native language of target group in recruitment materials, correspondence, and study materials + Latino; Spanish CTN MET study (Carroll et al., 2009; Suarez-Morales et al., 2007)
Community-based participatory research + American Indian/Alaska Native; substance dependent (The Southwest Node of NIDA CTN, Na'nizhoozhi Center, Inc., and Navajo Nation Human Research Review Board [2007], as cited in Burlew, Larios et al., 2011) + Black; women seeking faith-based intervention (Wingood, Simpson-Robinson, Braxton, & Raiford, 2011) + Racial/Ethnic Minorities; meta-analysis (De Las Nueces et al., 2012) + Black; male HIV prevention (Fortune, Wright, Juzang, & Bull, 2010)	Media resources: Letters, distributions of flyers sent, newspaper, TV/radio announcements, and/or other advertising about study + Latino; Spanish CTN MET study (Carroll et al., 2009; Suarez-Morales et al., 2007) + Black; smoking cessation (Okuyemi et al., 2007) + Black; lifestyle change and blood pressure control (Kennedy et al., 2010) − Black, Latino, & Asian; Panic Disorder (Mendoza, Williams, Chapman, & Powers, 2012) − Black & Latino; HIV positive prevention intervention (Hatfield et al., 2010)
Conducting recruitment or research in community settings (e.g. barber shops, beauty shops, or locations target group frequents) + Black; Black Barbershop Health Outreach Program (Releford, Frencher, & Yancey, 2010) + Black; women's health intervention (Johnson, Ralston, & Jones, 2010) + Latino, Black, & American Indian/Alaska Native; drug prevention program (Harachi, Catalano, & Hawkins, 1997)	Adding health information about the target group to the recruitment flyers, brochures, or letters − Racial/Ethnic minorities; worksite dietary intervention (Kiernan, Phillips, Fair, & King, 2000) Insertion of picture of the racial/ethnic minority researcher − Black; cancer prevention (Satia, Galanko, & Rimer, 2005) Including a culturally sensitive letter stressing the need for more information on the target group − Black; cancer prevention (Satia et al., 2005) Clinic referrals from existing referral sources − Black, Latino, & Asian; Panic Disorder (Mendoza et al., 2012) − Black; adolescents with ADHD and Substance Use Disorder (Holmes et al., n.d.) Increased number of recruitment sites − Black, Latino, & Asian; Panic Disorder (Mendoza et al., 2012)
Use of community resources in conjunction with word of mouth + Latino; behavioral observation study (Rodríguez et al., 2006)	Using recruiters from the same racial/ethnic group as target group − Black; psychiatric sample (Thompson, Neighbors, Munday, & Jackson, 1996) + Black; women with breast cancer (Moorman, Newman, Millikan, Tse, & Sandler, 1999) + Latino; women's health initiative (Larkey et al., 2002)

Note. +/− indicates whether the method improved or did not improve recruitment, respectively.

successful in increasing Black participants, Respondent-Driven Sampling was less effective among Latino participants (Hatfield et al., 2010), who are more likely to participate, for instance, when research teams use cultural adaptation techniques (e.g., recruiting participants in their native language; disseminating advertising and recruitment materials in the participants' native language; Carroll et al., 2009; Suarez-Morales et al., 2007). Similarly, the literature review revealed that while some recruitment methods might be successful for one group and not another, the efficacy of a recruit-

ment method could also vary within a racial/ethnic group. For example, ethnic matching of recruiters to the target group was found to be effective among a Black psychiatric sample (Thompson, Neighbors, Munday, & Jackson, 1996), yet less effective among a Black breast cancer sample (Moorman et al., 1999). We recommend that researchers examine existing literature for information on specific racial/ethnic minority groups or types (e.g., immigrants, homeless) to determine what recruitment methods might be effective among that participant population.

Step 5: Data Analysis

After the data are collected, several decisions about data analyses have important implications for the results. Sections 5A and 5B discuss two of these issues. Section 5A discusses the potential pitfalls of using standardized norms to interpret the findings. Section 5B discusses the advantages and disadvantages of combining subgroups for the analyses.

Step 5A: Consider the Hazards of Using Norms and Cutoff Scores

Despite the temptation to use published norms to interpret the scores of one's sample, we caution the researcher to be aware of potential unexpected consequences. In an earlier publication, Burlew et al. (2009) used the Minnesota Multiphasic Personality Inventory (MMPI)-2 to illustrate potential problems associated with using norms based on the scores of other groups. That illustration made several points. First, if only a few members of a specific racial/ethnic minority group are in the standardization sample, that smaller group's contribution to the overall mean may be negligible. In effect, the scores of the majority group are being used to interpret the scores of the smaller size group. Burlew et al. (2009) also walk through two examples that use standardized scores (e.g., largely based on the scores of the White respondents) to evaluate the scores for both an Asian American male and female on the psychasthenia (Pt) scale of the MMPI. Since Asian American males have higher mean scores than White males on the Pt, an Asian American male with a raw score of 21 would be in the clinical range if evaluated based on the predominately White standardization sample ($T = 66$) but in the normal range ($T = 60$) if evaluated against a predominately Asian American sample. The opposite happens for Asian American women. Since Asian American women have lower scores than White women on the PT scale, an Asian American woman with a raw score of 21 would be in the normal range if her scores were evaluated based on a predominately White sample ($T = 63$) but in the clinical range if evaluated based on other Asian American women ($T = 72$).

This example raises the question of whether separate norms should be used to evaluate various racial/ethnic minority groups. That question has been a source of considerable controversy. The argument for separate norms, as the above MMPI-2 example illustrates, is that separate norms may better reflect cultural, educational, socioeconomic, and other factors that may influence an individual's score. However, one argument against is that separate norms alone do not address the lack of cultural equivalence in the measure itself (Manly, 2005).

Until this issue is resolved, perhaps the best recommendation for those conducting research on racial/ethnic minorities might be to limit the use of measures that require the use of norms, to include plans to establish norms for the target group in their study design, or at least to include a disclaimer if a norm based on one group is used to evaluate a different target group.

Step 5B: Consider the Hazards of Combining Ethnic Groups for the Analyses

Samples with small numbers of several different racial/ethnic minority groups sometimes tempt researchers to combine racial/ethnic minority groups particularly for race-comparison studies. However, this strategy risks overlooking important racial/ethnic group differences. For example, in a study evaluating the effectiveness of a trauma-informed intervention, Amaro et al. (2007) found that a subgroup of Latinos had better outcomes (i.e., fewer symptoms of posttraumatic stress disorder) than the Black group. This difference likely would have gone unnoticed if the two groups had been combined.

One plausible but perhaps imperfect strategy is to compare the combined group of ethnic minorities to the White group as a first step. Then, if differences are evident, a potential next step might be to conduct additional analyses to identify the racial/ethnic minority groups that account for that difference. However, Figure 1 illustrates the argument against that approach.

The hypothetical example in Figure 1 includes four racial/ethnic minority groups and a White group. Note that the combined group looks similar to the White group. However, the scores of two of the racial/ethnic minority groups are above the White sample and the other two are below. If the first step were to look at the combined sample, the researcher may mistakenly conclude that the racial/ethnic minority groups are similar to the White group and not look further. Moreover, if the plan is to look at each subgroup as a second step even if the combined group appears similar to the White group, then one might wonder what purpose the initial step, comparing the combined sample to the White group, actually served.

Researchers sometimes combine racial/ethnic minority groups because they have small numbers. Hoyle (1999) provides constructive alternatives for addressing small samples (e.g., strategies for maximizing power, use of effect sizes) that may be an alternative to combining racial/ethnic minority groups. In addition to Hoyle's suggestions, we recommend that researchers take proactive steps to shape a study's sample composition when developing the recruitment plan. For example, a researcher might select recruitment sites where one might be able to enroll larger numbers of racial/ethnic minorities. Moreover, if recruiting sufficient numbers of several racial/ethnic minority groups is not feasible, the research team might focus on enrolling larger numbers of one specific racial/ethnic minority group in a larger study rather recruiting smaller numbers of multiple racial/ethnic minority group members.
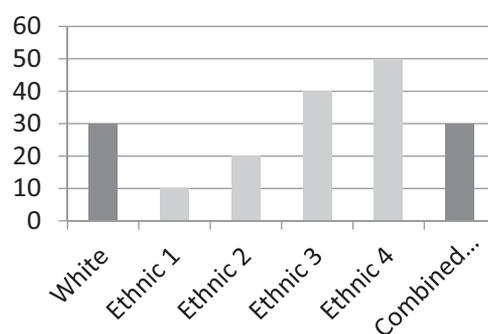

Figure 1. Illustration of combined ethnic groups.

Step 6: Dissemination

Dissemination includes the circulation of research to lay as well as professional (Schillinger, 2010). Publication in academic journals is the gold standard in legitimizing research. This section describes our recommendations for addressing pertinent issues facing researchers at the dissemination stage including selection of a journal, the benefits of a diversified portfolio, and dissemination to lay audiences.

Step 6A. Selecting an Appropriate Journal

Differing opinions regarding the value of race scholarship influence the opportunities for publishing work on racial/ethnic minorities (Nagayama Hall & Maramba, 2001). Scholars need to be aware of the expectations of their academic units for appropriate publication outlets, especially since the circulation rates of race/ethnicity journals often result in lower impact factors (IFs). Ideally, discussions with department heads about publishing should begin during hiring and contract negotiations but should be ongoing. We also encourage junior investigators to (a) seek advice from senior colleagues on appropriate publication venues and (b) collect information on specific racial/ethnic journals in the event it becomes necessary to educate department members or promotion and tenure committees on the typical rankings of race/ethnicity journals and citation counts of top scholars in race scholarship. For example, despite the fact that the acceptance rates for racial/ethnic journals mirror some of the top journals in their fields, the top 10 racial/ethnic journals have IFs ranging from zero (or absent) to 3.5, whereas top psychological journals exceed 12.85 (SCImago Journal and Country Rating [SJR], 2015). Moreover, mainly due to a restricted readership, Nagayama Hall and Maramba (2001) found that even the top scholars with the most publications in the field of race scholarship ranked low on citation counts, with a 6-year average of about 10 citations.

Step 6B. Diversify Your (Research) Portfolio

Scholars of color are often faced with the dilemma of pursuing mainstream versus race scholarship given the frequent devaluation of race research (Allen, Epps, Guillory, Suh, & Bonous-Hammarth, 2000; Padilla, 1994). Ultimately, the answer about whether to pursue mainstream or race/ethnicity research may not always be an either/or but can be both/and scenario. Scholars might consider developing a research portfolio that includes work that interests racial/ethnic journals (i.e., multicultural) as well as work attractive to mainstream journals (generalizability).

We suggest several strategies for increasing opportunities to publish in mainstream journals:

- **Assess**: Review the publications for the past few years to select a mainstream journal that has published cultural or race research.
- **Communicate**: Talk to journal editors about current research priorities before submitting an article.
- **Frame**: When appropriate, couch research submitted to a mainstream journal within a universal principal or a generalizability theory in the field (e.g., the application of Social-Exchange Theory to altruism in Blacks).

- **Coordinate**: Seek opportunities to publish articles on racial/ethnic minority groups in a special issue or to organize a special issue of a journal focusing on racial/ethnic topics.

Step 6C. Responsibly Disseminate to Lay Audiences

Researchers have an ethical responsibility to those who permit us to do research in their communities. *Helicopter researchers* extract information from the community and then disappear, whereas responsible researchers include the community from the planning through the dissemination stages. Examples of lay dissemination include publishing the findings in nonacademic outlets (e.g., blogs, news outlets, social media), presenting findings at community meetings and other venues (e.g., radio/news shows), and sharing results with key stakeholders including community organizers, policymakers, and social service agencies. We recommend broad and varied dissemination in order to increase the potential for both actual and meaningful scientific impact on diverse communities.

Discussion

The shift in representation of diverse groups in the United States coupled with health disparities that disproportionately affect these groups highlight the necessity for accurate research on these groups. Methodologically sound high-quality research has the potential to improve treatment and provide data for effective policies. Despite the consensus that sound research can contribute to achieving health equity, traditional theories, research methods, and dissemination approaches may be inadequate for research on racial/ethnic minorities. In a systematic process, the authors described common issues in health disparities research and provided readers with guidance on best practices to improving research methods at the planning, design, measurement, recruitment, data analysis, and dissemination steps.

Implications for Research, Policy and Practice

Despite the challenges of conducting valid research on racial/ethnic minorities, conducting culturally competent research has unique implications for researchers, policies, and practice. Mandates to include minorities (and women) in research provide little guidance on appropriate implementation. In order to generate impactful data on racial/ethnic minorities, future researchers who include people of color may consider implementing the strategies outlined herein. We invite researchers to consider, at each stage, where these alternative strategies may be appropriate to increase the knowledge base. Researchers might advance their careers by designing research programs consisting of sequential projects that address generalist, group difference, and multicultural research questions. Moreover, publication plans might include supplemental projects that investigate measurement equivalence and/or assess potential race-of-interviewer/race-of-coder effects on outcomes.

Results from research studies benefit society and should be accessible to all. The information has the ability to enhance the health of the population by generating knowledge and informing interventions. Many funding agencies have existing dissemination

policies (NIH, 2018; NSF, 2018). Policy developers may consider mandating dissemination practices that demonstrate evidence of reaching ethnic/racial minority groups and those serving these populations. We recommend that health disparities researchers, and/or those not federally funded, take extra efforts to ensure accessibility, such as publishing in open-access journals. Institutional Review Boards could also mandate that researchers provide a dissemination plan with their proposed research. This approach may help establish the benefits to research participants and increase the responsible conduct of research.

High-quality research findings should also inform clinical practice. However, practitioners are responsible for staying current on the literature through searching, continuing education, and attending conferences (Wysocki & Bookbinder, 2005). Health care licensing boards might consider mandating continuing education training with diverse populations. For example, the Ohio Board of Psychology mandates that psychologists complete 4 hours of training in ethics or cultural competence biannually (Ohio Psychological Association, 2018).

The present article makes a significant contribution to the field of health disparities research. The shift in representation of diverse groups in the United States coupled with the lagging research evidence highlights the necessity for culturally conscious research. Some researchers may have a training gap in culturally informed research design and implementation. This paper helps fill this gap by providing multiple strategies to make cultural considerations at each phase of a research project. These strategies can increase individual productivity, provide more research evidence for understudied populations, enhance disseminations practices, and inform intervention and prevention efforts. Implementing these strategies will contribute to closing the health gap.

Keywords: health disparities; health disparities research; cultural competence; culturally competent research

References

Allen, J., & Walsh, J. A. (2000). A construct-based approach to equivalence: Methodologies for cross-cultural/multicultural personality assessment research. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 63–85). Mahwah, NJ: Lawrence Erlbaum.

Allen, W., Epps, E., Guillory, E., Suh, S., & Bonous-Hammarth, M. (2000). The Black academic: Faculty status among African Americans in U.S. higher education. *The Journal of Negro Education, 69,* 112–127.

Alvarez, R. A., Vasquez, E., Mayorga, C. C., Feaster, D. J., & Mitrani, V. B. (2006). Increasing minority research participation through community organization outreach. *Western Journal of Nursing Research, 28,* 541–560. http://dx.doi.org/10.1177/0193945906287215

Amaro, H., Dai, J., Arévalo, S., Acevedo, A., Matsumoto, A., Nieves, R., & Prado, G. (2007). Effects of integrated trauma treatment on outcomes in a racially/ethnically diverse sample of women in urban community-based substance abuse treatment. *Journal of Urban Health, 84,* 508–522. http://dx.doi.org/10.1007/s11524-007-9160-z

American Psychological Association (APA). (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

American Psychological Association (APA). (2017). *Multicultural guidelines: An ecological approach to context, identity, and intersectionality.* Retrieved from http://www.apa.org/about/policy/multicultural-guidelines.pdf

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73,* 3–25. http://dx.doi.org/10.1037/amp0000191

Azibo, D. (1988). Understanding the proper and improper usage of the comparative research framework. *Journal of Black Psychology, 15,* 81–91. http://dx.doi.org/10.1177/00957984880151010

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black–White differences in response styles. *Public Opinion Quarterly, 48,* 491–509. http://dx.doi.org/10.1086/268845

Bhopal, R., & Donaldson, L. (1998). White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. *American Journal of Public Health, 88,* 1303–1307. http://dx.doi.org/10.2105/AJPH.88.9.1303

Bravo, M. (2003). Instrument development: Cultural adaptations in ethnic minority research. In G. Bernal, J. Trimble, A. Burlew, & F. Leong (Eds.), *Handbook of racial ethnic minority psychology* (pp. 220–236). Newbury Park, CA: SAGE. http://dx.doi.org/10.4135/9781412976008.n12

Brislin, R. W. (1993). *Understanding culture's influence on behavior.* New York, NY: Harcourt Brace Jovanovich.

Brody, G. H., Murry, V. M., Gerrard, M., Gibbons, F. X., Molgaard, V., McNair, L., . . . Neubaum-Carlan, E. (2004). The Strong African American Families Program: Translating research into prevention programming. *Child Development, 75,* 900–917. http://dx.doi.org/10.1111/j.1467-8624.2004.00713.x

Brown, G., Marshall, M., Bower, P., Woodham, A., & Waheed, W. (2014). Barriers to recruiting ethnic minorities to mental health research: A systematic review. *International Journal of Methods in Psychiatric Research, 23,* 36–48. http://dx.doi.org/10.1002/mpr.1434

Bureau of Labor Statistics. (1995). *Preference for racial or ethnic terminology.* Retrieved from http://www.infoplease.com/ipa/A0762158.html

Burlew, A. K., Feaster, D., Brecht, M. L., & Hubbard, R. (2009). Measurement and data analysis in research addressing health disparities in substance abuse. *Journal of Substance Abuse Treatment, 36,* 25–43. http://dx.doi.org/10.1016/j.jsat.2008.04.003

Burlew, A. K., Larios, S., Suarez-Morales, L., Holmes, B., Venner, K., & Chavez, R. (2011). Increasing ethnic minority participation in substance abuse clinical trials: Lessons learned in the National Institute on Drug Abuse's Clinical Trials Network. *Cultural Diversity & Ethnic Minority Psychology, 17,* 345–356. http://dx.doi.org/10.1037/a0025668

Burlew, A. K., Weekes, J. C., Montgomery, L., Feaster, D. J., Robbins, M. S., Rosa, C. L., . . . Wu, L.-T. (2011). Conducting research with racial/ethnic minorities: Methodological lessons from the NIDA Clinical Trials Network. *The American Journal of Drug and Alcohol Abuse, 37,* 324–332. http://dx.doi.org/10.3109/00952990.2011.596973

Carroll, K. M., Martino, S., Ball, S. A., Nich, C., Frankforter, T., Anez, L. M., . . . Farentinos, C. (2009). A multisite randomized effectiveness trial of motivational enhancement therapy for Spanish-speaking substance users. *Journal of Consulting and Clinical Psychology, 77,* 993–999. http://dx.doi.org/10.1037/a0016489

Carvalho, J. O., Tommet, D., Crane, P. K., Thomas, M. L., Claxton, A., Habeck, C., . . . Romero, H. R. (2015). Deconstructing racial differences: The effects of quality of education and cerebrovascular risk factors. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 70,* 545–556. http://dx.doi.org/10.1093/geronb/gbu086

Census Briefs. (2012, March 12). *The Asian population: 2010.* Retrieved from http://www.census.gov/prod/cen2010/briefs/c2010br-11.pdf

Cornell, S., & Hartmann, D. (1998). *Ethnicity and race: Making identities in a changing world.* Thousand Oaks, CA: Pine Forge Press.

Davis, D. W. (1997). The direction of race of interviewer effects among African-Americans: Donning the Black mask. *American Journal of Political Science, 41,* 309–322. http://dx.doi.org/10.2307/2111718

Davis, D. W., & Silver, B. D. (2003). Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science, 47,* 33–45. http://dx.doi.org/10.1111/1540-5907.00003

De Las Nueces, D., Hacker, K., DiGirolamo, A., & Hicks, L. S. (2012). A systematic review of community-based participatory research to enhance clinical trials in racial and ethnic minority groups. *Health Services Research, 47,* 1363–1386. http://dx.doi.org/10.1111/j.1475-6773.2012.01386.x

Douglas, S., & Craig, S. S. (2007). Collaborative and iterative translation: An alternative approach to instrument translation. *Journal of International Marketing, 15,* 30–43. http://dx.doi.org/10.1509/jimk.15.1.030

Fernández, M. I., Varga, L. M., Perrino, T., Collazo, J. B., Subiaul, F., Rehbein, A., . . . Bowen, G. S. (2004). The Internet as recruitment tool for HIV studies: Viable strategy for reaching at-risk Hispanic MSM in Miami? *AIDS Care, 16,* 953–963. http://dx.doi.org/10.1080/09540120412331292480

Fernando, S. (2010). *Mental health, race and culture.* New York, NY: Palgrave Macmillan. http://dx.doi.org/10.1007/978-1-137-01368-2

Fly, E. (2010, September 10). 'Mulatto': Racial slur or socially acceptable. *Parlour Magazine.* Retrieved from http://parlourmagazine.com/2010/09/mulatto-racial-slur-or-socially-exceptable/

Fortune, T., Wright, E., Juzang, I., & Bull, S. (2010). Recruitment, enrollment and retention of young black men for HIV prevention research: Experiences from the 411 for Safe Text project. *Contemporary Clinical Trials, 31,* 151–156. http://dx.doi.org/10.1016/j.cct.2009.12.004

Ghee, A. C., Johnson, C. S., & Burlew, A. K. (2010). Measurement equivalence of the Trauma Symptom Checklist-40 for chemically dependent African American and Caucasian women: A preliminary study. *Journal of Aggression, Maltreatment & Trauma, 19,* 820–838. http://dx.doi.org/10.1080/10926771.2010.522949

Gonzales, N. A., Cauce, A. M., & Mason, C. A. (1996). Interobserver agreement in the assessment of parental behavior and parent–adolescent conflict: African American mothers, daughters, and independent observers. *Child Development, 67,* 1483–1498. http://dx.doi.org/10.2307/1131713

Graham, S. (1992). "Most of the subjects were White and middle class": Trends in published research on African Americans in selected APA journals, 1970–1989. *American Psychologist, 47,* 629–639. http://dx.doi.org/10.1037/0003-066X.47.5.629

Guerra, N. G., & Jagers, R. (1998). The importance of culture in the assessment of children and youth. In V. McLoyd & L. Steinberg (Eds.), *Studying minority adolescents: Conceptual, methodological, and theoretical issues* (pp. 167–181). Mahwah, NJ: Erlbaum.

Hall, G. C. N., DeGarmo, D. S., Eap, S., Teten, A. L., & Sue, S. (2006). Initiation, desistance, and persistence of men's sexual coercion. *Journal of Consulting and Clinical Psychology, 74,* 732–742. http://dx.doi.org/10.1037/0022-006X.74.4.732

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

Harachi, T. W., Catalano, R. F., & Hawkins, J. D. (1997). Effective recruitment for parenting programs within ethnic minority communities. *Child & Adolescent Social Work Journal, 14,* 23–39. http://dx.doi.org/10.1023/A:1024540829739

Hartlieb, K. B., Jacques-Tiura, A. J., Naar-King, S., Ellis, D. A., Jen, K. L., & Marshall, S. (2015). Recruitment strategies and the retention of obese urban racial/ethnic minority adolescents in clinical trials: The FIT families project, MI, 2010–2014. *Preventing Chronic Disease, 12,* 140409. http://dx.doi.org/10.5888/pcd12.140409

Hatfield, L. A., Ghiselli, M. E., Jacoby, S. M., Cain-Nielsen, A., Kilian, G., McKay, T., & Rosser, B. R. (2010). Methods for recruiting men of color who have sex with men in prevention-for-positives interventions. *Prevention Science, 11,* 56–66. http://dx.doi.org/10.1007/s11121-009-0149-6

Helms, J. E., Jernigan, M., & Mascher, J. (2005). The meaning of race in psychology and how to change it: A methodological perspective. *American Psychologist, 60,* 27–36. http://dx.doi.org/10.1037/0003-066X.60.1.27

Holmes, B., Pressley, K., Haynes, L., Tyson, C., & Riggs, P. (n.d.). *Using respondent-driven sampling (RDS) to enhance recruitment of dually-diagnosed adolescent clinical trial participants.* Retrieved from http://ctndisseminationlibrary.org/PDF/374.pdf

Hoyle, R. (1999). *Statistical strategies for small sample research.* Thousands Oaks, CA: SAGE.

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

Johnson, L. T., Ralston, P. A., & Jones, E. (2010). Beauty salon health intervention increases fruit and vegetable consumption in African-American women. *Journal of the American Dietetic Association, 110,* 941–945. http://dx.doi.org/10.1016/j.jada.2010.03.012

Johnson, T. P., & Parsons, J. A. (1994). Interviewer effects on self-reported substance use among homeless persons. *Addictive Behaviors, 19,* 83–93. http://dx.doi.org/10.1016/0306-4603(94)90054-X

Joireman, J., & Van Lange, P. (2015). *How to publish high-quality research.* Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/14525-000

Kaufman, S. B., & Wicherts, J. (2011, May 21). Black women are not (rated) less attractive! Our independent analysis of the Add Health dataset. *Psychology Today.* Retrieved from https://www.psychologytoday.com/blog/beautiful-minds/201105/black-women-are-not-rated-less-attractive-our-independent-analysis-the

Kennedy, B. M., Kumanyika, S., Ard, J. D., Reams, P., Johnson, C. A., Karanja, N., . . . Harsha, D. W. (2010). Overall and minority-focused recruitment strategies in the PREMIER multicenter trial of lifestyle interventions for blood pressure control. *Contemporary Clinical Trials, 31,* 49–54. http://dx.doi.org/10.1016/j.cct.2009.10.002

Kiernan, M., Phillips, K., Fair, J. M., & King, A. C. (2000). Using direct mail to recruit Hispanic adults into a dietary intervention: An experimental study. *Annals of Behavioral Medicine, 22,* 89–93. http://dx.doi.org/10.1007/BF02895172

Kim, G., DeCoster, J., Bryant, A. N., & Ford, K. L. (2016). Measurement equivalence of the K6 Scale: The effects of race/ethnicity and language. *Assessment, 23,* 758–768. http://dx.doi.org/10.1177/1073191115599639

Kim, K. H., & Zabelina, D. (2015). Cultural bias in assessment: Can creativity assessment help? *The International Journal of Critical Pedagogy, 6,* 129–148.

Knight, G. P., Little, M., Losoya, S. H., & Mulvey, E. P. (2004). The self-report of offending among serious juvenile offenders: Cross-gender, cross-ethnic/race measurement equivalence. *Youth Violence and Juvenile Justice, 2,* 273–295. http://dx.doi.org/10.1177/1541204004265878

Larkey, L. K., Staten, L. K., Ritenbaugh, C., Hall, R. A., Buller, D. B., Bassford, T., & Altimari, B. R. (2002). Recruitment of Hispanic women to the Women's Health Initiative. The case of embajadoras in Arizona. *Controlled Clinical Trials, 23,* 289–298. http://dx.doi.org/10.1016/S0197-2456(02)00190-3

Maccoby, E., & Maccoby, N. (1954). The interview: A tool of social science. In L. Gardiner (Ed.), *Handbook of social psychology* (Vol. 1, pp. 449–487). Cambridge, MA: Addison Wesley.

Manly, J. J. (2005). Advantages and disadvantages of separate norms for African Americans. *The Clinical Neuropsychologist, 19,* 270–275. http://dx.doi.org/10.1080/13854040590945346

McHorney, C. A., & Fleishman, J. A. (2006). Assessing and understanding measurement equivalence in health outcome measures. Issues for further quantitative and qualitative inquiry. *Medical Care, 44,* S205–S210. http://dx.doi.org/10.1097/01.mlr.0000245451.67862.57

Melby, J. N., Hoyt, W. T., & Bryant, C. M. (2003). A generalizability approach to assessing the effects of ethnicity and training on observer ratings of family interactions. *Journal of Social and Personal Relationships, 20,* 171–191. http://dx.doi.org/10.1177/02654075030202003

Mendoza, D. B., Williams, M. T., Chapman, L. K., & Powers, M. (2012). Minority inclusion in randomized clinical trials of panic disorder. *Journal of Anxiety Disorders, 26,* 574–582. http://dx.doi.org/10.1016/j.janxdis.2012.02.011

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44,* S69–S77. http://dx.doi.org/10.1097/01.mlr.0000245438.73837.89

Moorman, P. G., Newman, B., Millikan, R. C., Tse, C. K., & Sandler, D. P. (1999). Participation rates in a case-control study: The impact of age, race, and race of interviewer. *Annals of Epidemiology, 9,* 188–195. http://dx.doi.org/10.1016/S1047-2797(98)00057-X

Nagayama Hall, G. C., & Maramba, G. G. (2001). In search of cultural diversity: Recent literature in cross-cultural and ethnic minority psychology. *Cultural Diversity & Ethnic Minority Psychology, 7,* 12–26. http://dx.doi.org/10.1037/1099-9809.7.1.12

Nagayama Hall, G. C., Teten, A. L., DeGarmo, D. S., Sue, S., & Stephens, K. A. (2005). Ethnicity, culture, and sexual aggression: Risk and protective factors. *Journal of Consulting and Clinical Psychology, 73,* 830–840. http://dx.doi.org/10.1037/0022-006X.73.5.830

National Institutes of Health (NIH). (2001). *Inclusion of women and minorities as participants in research involving human subjects—Policy implementation.* Retrieved from http://grants.nih.gov/grants/funding/women_min/women_min.htm

National Institutes of Health (NIH). (2016). *NIH-wide strategic plan.* Retrieved from https://www.nih.gov/sites/default/files/about-nih/strategic-plan-fy2016-2020-508.pdf

National Institutes of Health (NIH). (2017). *Amendment: NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research.* Retrieved from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-014.html

National Institutes of Health (NIH). (2018). *Research portfolio online reporting tools (RePORT).* Retrieved from https://report.nih.gov/

National Research Council (NRC). (2004). Measuring race, ethnicity, socioeconomic position, and acculturation. In M. Ver Ploeg & E. Perrin (Eds.), *Eliminating health disparities: Measurement and data needs* (pp. 41–61). Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/read/10979/chap./5

National Science Foundation (NSF). (2018). *NSF Public Access Repository (NSF-PAR).* Retrieved from https://par.nsf.gov/

Navajo Nation Human Research Review Board. (2007). *Procedural guidelines for principal investigators.* Window Rock, AZ: Navajo Division of Health.

Ohio Psychological Association (OPA). (2018). *OPA-MCE tracking information for psychologists.* Retrieved from http://ohpsych.site-ym.com/page/CETrackingInfo

Okuyemi, K. S., Cox, L. S., Nollen, N. L., Snow, T. M., Kaur, H., Choi, W., . . . Ahluwalia, J. S. (2007). Baseline characteristics and recruitment strategies in a randomized clinical trial of African-American light smokers. *American Journal of Health Promotion, 21,* 183–191. http://dx.doi.org/10.4278/0890-1171-21.3.183

Padilla, A. (1994). Ethnic minority scholars, research, and mentoring: Current and future issues. *Educational Researcher, 23,* 24–27.

Peteet, B. J., Montgomery, L., & Weekes, J. C. (2015). Predictors of imposter phenomenon among talented ethnic minority undergraduate students. *The Journal of Negro Education, 84,* 175–186. http://dx.doi.org/10.7709/jnegroeducation.84.2.0175

Releford, B. J., Frencher, S. K., Jr., & Yancey, A. K. (2010). Health promotion in barbershops: Balancing outreach and research in African American communities. *Ethnicity & Disease, 20,* 185–188.

Robinson, W. T., Risser, J. M., McGoy, S., Becker, A. B., Rehman, H., Jefferson, M., . . . Tortu, S. (2006). Recruiting injection drug users: A three-site comparison of results and experiences with respondent-driven and targeted sampling procedures. *Journal of Urban Health, 83*(Suppl. 1), 29–38. http://dx.doi.org/10.1007/s11524-006-9100-3

Rodríguez, M. D., Rodríguez, J., & Davis, M. (2006). Recruitment of first-generation Latinos in a rural community: The essential nature of personal contact. *Family Process, 45,* 87–100. http://dx.doi.org/10.1111/j.1545-5300.2006.00082.x

Sandefur, G. D., Campbell, M. E., & Eggerling-Boeck, J. (2004). Racial and ethnic identification, official classifications, and health disparities. In N. B. Anderson, R. A. Bulatao, & B. Cohen (Eds.), *Critical perspectives on racial and ethnic differences in health in late life* (pp. 25–52). Washington, DC: National Academies Press. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK25522/

Satia, J. A., Galanko, J. A., & Rimer, B. K. (2005). Methods and strategies to recruit African Americans into cancer prevention surveillance studies. *Cancer Epidemiology, Biomarkers & Prevention, 14,* 718–721. http://dx.doi.org/10.1158/1055-9965.EPI-04-0132

Schillinger, D. (2010). *An introduction to effectiveness, dissemination and implementation research.* San Francisco, CA: Clinical Translational Science Institute, University of California San Francisco. Retrieved from https://accelerate.ucsf.edu/files/CE/edi_introguide.pdf

SCImago Journal and Country Rating (SJR). (2016). *Journal rankings.* Retrieved from http://www.scimagojr.com/journalrank.php?country=U.S

Sigelman, L., Tuch, S. A., & Martin, J. K. (2005). What's in a name? Preference for "Black" versus "African-American" among Americans of African descent. *Public Opinion Quarterly, 69,* 429–438. http://dx.doi.org/10.1093/poq/nfi026

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology, 69,* 797.

Stokes, L., & Yeh, M. Y. (1988). Searching for causes of interviewer effects in telephone surveys. In R. M. Groves, P. P. Biememr, L. E. Lyberg, J. T. Massey, W. L. Nicholls, II, & J. J. Waksberg (Eds.), *Telephone survey methodology* (pp. 357–383). New York, NY: Wiley.

Suarez-Morales, L., Matthews, J., Martino, S., Ball, S. A., Rosa, C., Farentinos, C., . . . Carroll, K. M. (2007). Issues in designing and implementing a Spanish-language multi-site clinical trial. *The American Journal on Addictions, 16,* 206–215. http://dx.doi.org/10.1080/10550490701375707

Substance Abuse and Mental Health Services Administration (SAMHSA). (2014). *Introduction to cultural competence* (Treatment Improvement Protocol (TIP) series, No. 59). Rockville, MD: Center for Substance Abuse Treatment.Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK248431/

Sue, S., & Dhindsa, M. K. (2006). Ethnic and racial health disparities research: Issues and problems. *Health Education & Behavior, 33,* 459–469. http://dx.doi.org/10.1177/1090198106287922

Taylor, P., Lopez, M. H., Martínez, J. H., & Velasco, G. (2012). When labels don't fit: Hispanics and their views of identity. *The Pew Hispanic Center.* Retrieved from http://m.fronterasdesk.org/sites/default/files/field/docs/2012/04/04/Hispanic-Identity.pdf

Teresi, J. A., Stewart, A. L., Morales, L. S., & Stahl, S. M. (2006). Measurement in a multi-ethnic society. Overview to the special issue. *Medical Care, 44,* S3–S4. http://dx.doi.org/10.1097/01.mlr.0000245437.46695.4a

Thompson, E. E., Neighbors, H. W., Munday, C., & Jackson, J. S. (1996). Recruitment and retention of African American patients for clinical research: An exploration of response rates in an urban psychiatric

hospital. *Journal of Consulting and Clinical Psychology, 64,* 861–867. http://dx.doi.org/10.1037/0022-006X.64.5.861

Tuliao, A. P., Landoy, B. V., & McChargue, D. (2016). Factor structure and invariance test of the alcohol use disorder identification test (AU-DIT): Comparison and further validation in a U.S. and Philippines college student sample. *Journal of Ethnicity in Substance Abuse, 15,* 127–143. http://dx.doi.org/10.1080/15332640.2015.1011731

U.S. Bureau of the Census. (2001). *The two or more races population: 2000.* Retrieved from https://www.census.gov/prod/2001pubs/c2kbr01-6.pdf

U.S. Bureau of the Census. (2010). *The American Community Survey.* Retrieved from http://www2.census.gov/programs-surveys/acs/methodology/questionnaires/2010/quest10.pdf

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4–70. http://dx.doi.org/10.1177/109442810031002

Waheed, W., Hughes-Morley, A., Woodham, A., Allen, G., & Bower, P. (2015). Overcoming barriers to recruiting ethnic minorities to mental health research: A typology of recruitment strategies. *BMC Psychiatry, 15,* 101. http://dx.doi.org/10.1186/s12888-015-0484-z

Waters, M. (1999). *Black identities: West Indian immigrant dreams and American realities.* New York, NY: Russell Sage Foundation.

Weaver, A., Himle, J. A., Taylor, R. J., Matusko, N. N., & Abelson, J. M. (2015). Urban vs. rural residence and the prevalence of depression and mood disorder among African American women and non-Hispanic White women. *Journal of the American Medical Association Psychiatry, 72,* 576–583. http://dx.doi.org/10.1001/jamapsychiatry.2015.10

Wingood, G. M., Simpson-Robinson, L., Braxton, N. D., & Raiford, J. L. (2011). Design of a faith-based HIV intervention: Successful collaboration between a university and a church. *Health Promotion Practice, 12,* 823–831. http://dx.doi.org/10.1177/1524839910372039

Wu, L. T., Pan, J. J., Blazer, D. G., Tai, B., Stitzer, M. L., & Woody, G. E. (2010). Using a latent variable approach to inform gender and racial/ethnic differences in cocaine dependence: A National Drug Abuse Treatment Clinical Trials Network study. *Journal of Substance Abuse Treatment, 38*(Suppl. 1), S70–S79. http://dx.doi.org/10.1016/j.jsat.2009.12.011

Wysocki, A., & Bookbinder, M. (2005). Implementing clinical practice changes: A practical approach. *Home Health Care Management & Practice, 17,* 164–174. http://dx.doi.org/10.1177/1084822304271790

Yancey, A. K., Ortega, A. N., & Kumanyika, S. K. (2006). Effective recruitment and retention of minority research participants. *Annual Review of Public Health, 27,* 1–28. http://dx.doi.org/10.1146/annurev.publhealth.27.021405.102113

Yoon, E., Chang, C. T., Kim, S., Clawson, A., Cleary, S. E., Hansen, M., . . . Gomes, A. M. (2013). A meta-analysis of acculturation/enculturation and mental health. *Journal of Counseling Psychology, 60,* 15–30. http://dx.doi.org/10.1037/a0030652