

An examination of the evidence in culturally adapted evidence-based or empirically supported interventions

Janet E. Helms

Boston College

Abstract

Measurement yields perhaps the most critical evidence influencing whether culturally adapted evidence-based practice (EBP) and empirically supported treatments (EST) are deemed more effective for African Americans, Latino/a Americans, Asian/Pacific Islander Americans, Native Americans, and related immigrant groups than standard treatments, as well as for determining the validity of results of surveys of health conditions in nondominant populations internationally. However, little attention has been given to measuring the effects of race and ethnic culture, as experiential constructs rather than sociodemographic categories, on diagnosis, the treatment process, and outcomes. Three meta-analyses of culturally adapted treatments and three studies cited in them were analyzed to determine the ways in which researchers incorporated measurement of racial and ethnic cultural dynamics as explicit factors in any phase of their interventions. The analysis revealed that researchers did not report adapting standard measures to address cultural influences, nor did they define symptoms from participants' cultural or racial experiences. The author concludes that although there are criteria for judging good research designs, which may or not be feasible for research on nondominant racial and ethnic groups, there are no paradigms for developing measures or for interpreting existing measures to incorporate ethnicity and racialized experiences. Some principles from cross-cultural assessment research (i.e., functional, conceptual, metric, and linguistic equivalence) are adapted to suggest how measures for investigating the effectiveness of culturally adapted interventions for nondominant ethnic and racialized groups might be developed and/or used more appropriately throughout the course of the intervention.

Keywords

cultural equivalence, culturally adapted interventions, culture, evidence-based practice, measurement, race

Corresponding author:

Janet E. Helms, Boston College, 317 Campion Hall, LSOE, Chestnut Hill, MA 02467, USA.

Email: Janet.Helms@bc.edu

In the United States (US) and internationally, researchers have begun to question the effectiveness of evidence-based practice (EBP) and empirically supported treatments (EST) for addressing the mental health concerns of inadequately served racial and ethnic cultural groups for whom standard treatments typically yield disparate outcomes relative to White service seekers (Liu et al., 2012; Office of the Surgeon General, 2001). In quantitative research on culture and treatment effectiveness, the emphasis in the US has been on mental health outcomes (e.g., alleviation of depression), whereas the international emphasis has tended toward physical health-related outcomes (e.g., obesity). Efforts to reduce racial/cultural health care disparities have resulted in treatment comparison studies in which the efficacy and effectiveness of culturally adapted therapies and psychoeducational interventions are compared to those of standard treatments by means of traditional research designs and measures. Most of these studies have not collected information on culture and race as separate sets of socialization experiences, each of which may have psychological consequences for service providers, service recipients, and their interactions.

The present paper examines the ways in which racial/ethnic cultural factors have been incorporated into the evidence used to assess the relative effectiveness of empirically supported (ES)/evidenced-based (EB) mental health treatments to make some recommendations for how conceptualizations of “evidence” in culturally adapted treatment studies ought to be modified to ensure that the racial/ethnic cultural experiences of service recipients from underserved groups are respected throughout the process. I critique the American Psychological Association’s (APA) perspective on culture and race in evidence-based practice (APA Presidential Task Force on Evidence-Based Practice, 2006), and studies based on that perspective, to argue that producing trustworthy evidence requires increased focus on service providers’ (e.g., therapists) and service seekers’ racial/cultural interactions with treatments. Although racial and ethnic concepts are addressed primarily from a U.S. perspective in this article, aspects of the argument pertain to international evidence-based interventions as well.

Conceptualizations of culture and race

Definitions of race and/or ethnic culture generally have specified broad demographic categories with the nature, number, and labels of the categories varying according to the historical and sociopolitical racial/ethnic-cultural composition of specific countries. The US uses more categories and labels for designating race and/or ethnic culture and more complex parameters for specifying who is a member of which group than some more homogeneous countries such as the United Kingdom (UK) where 90% of the population reportedly is labeled “White” or “European” (Bhopal et al., 2004, p. 78). In much of the international literature, demographic minorities are conceptualized as originating from different countries, regardless of their generational status in the country in which they reside, whereas in the US, a variety of criteria are considered in racialized groupings

including, but not limited to Spanish-language heritage, presumed African descent, skin color, country of origin, and social policy (Agyemang, Akins, & Bophal, 2012; Bhopal, 1997).

Moreover, world-wide, studies of evidence-based interventions have reified their countries' racial/ethnic categories by proceeding as if these categories caused treatment-related behavioral outcomes rather than seeking to identify salient life and socialization experiences associated with ascribed membership in labeled categories that might cause differential outcomes (Helms, Jernigan, & Mascher, 2005; Kirmayer, 2012). Thus, for the most part, in quantitative research, evidence of cultural or racial treatment effectiveness has been inferred from between-treatment-group differences in outcomes, usually within racial or ethnic cultural groups, without regard to whether the measures of outcomes are culturally congruent or meaningful to the groups.

Consistent with Helms and Cook's (1999) recommendations, I differentiate ethnic culture from race, where *ethnic culture* or *ethnicity* is defined as the values, beliefs, customs, and traditions transmitted from generation to generation by important socializing agents (e.g., parents, schools) in a person's life and internalized or learned by the person as a member of the ethnic cultural group; *ethnic group* refers to the kinship groups who practice the ethnic culture and may or may not share phenotypic traits (e.g., Latinas/os in the US). *Race* (actually "socio-race") is a factitious label assigned to individuals on the basis of presumed phenotypic characteristics (e.g., skin color) of the person or the person's ancestors. Race or racial categories do not signify specific behaviors, but people do acquire behaviors that often are assumed to be caused by race rather than by the racial socialization (e.g., racism) accorded them because they are perceived as belonging to particular groups that are differentially valued or devalued in society (Helms & Jernigan, et al. 2005).

The foregoing definitions imply that researchers ought to define and measure evidence in terms of what individuals have learned or acquired from their ethnic cultural or racial experiences rather than focusing on their sociodemographic designations as reflected in outward appearance or nativity status. In culturally adapted EBP, it is some psychological or behavioral aspect of the person's ethnicity, nativity status, or cultural health-belief system that the practitioner is seeking to engage or change.

Culture and race in evidence-based practice

The American Psychological Association (APA) defined *evidence-based practice* (EBP) in psychology as "the integration of the best available research with clinical expertise in the context of patient characteristics, *culture*, and preferences" (emphasis added; 2006, p. 284). This definition was so vague that it immediately generated considerable discussion and controversy about every aspect of it despite the efforts of the specially appointed APA Presidential Task Force on Evidence-Based Practice (henceforth the "Task Force," 2006) to clarify its meaning. Thus, in

the domain of “standard” therapy or treatment, scholars and practitioners debated matters including: (a) whether empirically supported treatments, evidence-based practice, and manualized treatments are synonymous (Gambrill, 2006; Wachtel, 2010); (b) what constitutes “clinical expertise” (Dimidjian & Hollon, 2011); and, to a lesser extent, (c) the correct interpretation of “patient characteristics, culture, and preferences.” Analogous critiques and debates have occurred specifically about the “culture” aspects of the APA definition (Cardemil, 2008; Castro, Barrera, & Streiker, 2010; Kirmayer, 2012; La Roche & Christopher, 2008; Roysicar, 2009), with added focus on questioning whether standard treatments should be modified to better reflect participants’ cultural backgrounds and/or whether alternative culturally responsive evidence-based interventions are needed.

By far, most of the critiques and research concerning the effectiveness of standard evidence-based practice/empirically supported therapies (EBP/EST) and culturally responsive interventions (CRI) in the US have investigated whether particular kinds of interventions yield desired outcomes within or across racial/ethnic groups. However, the influences of culture and race on the nature of help seekers’ responses to outcome measures and provider–recipient interactions have been virtually ignored. The failure to measure providers’ and help-seekers’ cultural-racial dynamics from treatment entry to outcome suggests that measurement of such factors is unimportant when, in fact, it is critical for demonstrating the benefits of CRIs relative to standard EBP/ESTs for African Americans, Latina/o Americans, Asian and Pacific Islander Americans, Native Americans (ALANAs), and related immigrant or ethnic cultural groups (RIGs). Standard EBP/ESTs have been indicted for disregarding critical life experiences of ALANA/RIGs and thereby potentially contributing to harmful outcomes for them (Helms & Cook, 1999; Kirmayer, 2012). Yet without integrating service recipients’ and providers’ cultural and racial life experiences or their respective interpretations of the condition for which service is sought into measures of their reactions to each other and to the treatment, neither harm nor helpfulness can be determined.

Furthermore, although the Task Force mentioned that awareness of “race,” as a facet of “the individual, social, and cultural context[s] of the patient,” is an aspect of clinical expertise (2006, p. 277), it relegated race to the help-seekers’ context. The Task Force did not acknowledge racial socialization (e.g., racism) as an acquired aspect of both the provider/researcher and help-seeking person, with implications for the treatment process, and, therefore, for evidence of treatment efficacy. This issue has also not been addressed in the growing literature on cultural adaptations of therapy or international health studies (Bhopal, 2013; Mir et al., 2012).

Why race and culture need special attention in evidence-based research

Most studies of traditional EBP/EST in the US, to the extent that they address race/culture at all, take their lead from the National Institutes of Health’s (NIH, 2001) policy and guidelines, which virtually mandated the inclusion of ALANA/

RIGs in samples, although there was no requirement that they or their cultures be studied in any depth. For their part, culturally responsive researchers derive their justification for studying cultural factors from U.S. Census demographers' projections that people of Color (i.e., all but single-race, non-Hispanic Whites) will constitute 57% of the U.S. population by 2060 (U.S. Census Bureau, 2012) and, thus, will require greater access to culturally responsive health care services. Neither of these two rationales attends to the limitations of sample size as an argument against research as usual, nor do they focus on the effects of different kinds of racial and cultural socialization on the treatment process generally and, therefore, on outcome measures. But culturally responsive and race-conscious researchers in the US have begun to focus on developing measures that reflect racial socialization of potential clients (Carter, Mazzula, Victoria, Vazquez, & Hall, 2013), cultural styles (Yeh, 2012), or therapists' self-evaluations of their own cultural competence (Sheu & Lent, 2007). Moreover, in the UK, researchers have offered guidelines for maximizing cultural validity of health survey questionnaires (Bhopal et al., 2004).

Sample issues

Even if Census Bureau projections of increased racial and ethnic diversity of the U.S. population are validated in the not too distant future, Whites will remain the largest single socioracial group and will continue to play a dominant role in defining the nature and norms of traditional measures of treatment process and outcomes. Consequently, researchers will still find themselves in situations in which the numbers of people representing a single socioracial group of color (e.g., Blacks, Asians/Pacific Islanders) or ethnic group (e.g., Pilipino Americans) may be so small that it is impossible to conduct psychometric analyses, develop measures, or analyze results obtained from them by means of gatekeeper-approved research strategies (Helms, Henze, Satiani, & Mascher, 2005; Mir et al., 2012). Researchers typically have addressed sample size concerns by aggregating across native and ethnic or immigrant samples and labeling such groups according to a shared socioracial designation. It may be useful to examine two such designations, "African Americans" and "Asian Americans/Pacific Islanders," to discuss why such aggregation is problematic for obtaining meaningful assessment and outcome data; similar issues are relevant for each of the ALANA/RIG groups and perhaps other populations internationally.

African Americans. Samples labeled "African American" typically include a variety of participants whose ethnic group origins are not indigenous African American (e.g., Haitians, Jamaicans, and Nigerians). In some sense researchers may be forgiven for not recognizing the importance of differentiating other African-descent groups from indigenous African Americans because government demographic statistics rarely acknowledge the existence of Black American ethnic groups, although the same cannot be said of the other ALANA socioracial groups. Extrapolating from Kent's (2007) analysis of the voluntary immigration history of the Black population in the US, it seems reasonable to define indigenous African

Americans as those individuals of some presumed African origins, who can trace their ancestry to at least one parent of alleged African descent who resided in the US prior to 1965, after which immigration laws began to change (e.g., the 1965 Immigration and Nationality Act Amendments) to allow the voluntary immigration of people of Color.

It is difficult to estimate the percentage of the African American/Black population that is not native because African-descent ethnic groups are not counted consistently and many self-identify as “White” even though they would not be perceived as such by many researchers or clinicians in the US. However, Kent reports that about 8% of the Black American population was born in countries other than the US. At least 10 African countries, the Caribbean, and various European countries account for this within-group ethnic diversity.

Many of the newly arrived ethnic groups reside in own-group ethnic communities separated from African Americans and White Americans. The separation allows these immigrant communities to maintain cultural practices from their countries of origin and shields them, to some extent, from the effects of a history of racial oppression as it has occurred in the US (though not necessarily from racial tensions in their countries of origin). Thus, researchers’ aggregated Black samples may be combinations of people from ethnic groups that do and do not define themselves as “African American,” have not had the same experiences of racism, do not all speak a U.S. American English dialect as a first language, and do not share the same cultural beliefs about mental health and health or exhibit the same level of trust or distrust in the mental health care system.

Asian Americans/Pacific Islanders (AAPIs). As one of the smaller socioracial groups, AAPIs comprise about 4.3% (13.8 million) of the U.S. population (Yeh, 2012). In describing the diversity of the AAPI socioracial group, Yeh noted that most AAPIs (69%) were born in countries other than the US, trace their ethnic origins to more than 20 countries, and consist of more than 66 “documented” ethnicities; moreover, more than 79% speak a language other than English at home. Such diversity means that much cultural variation exists among the immigrant AAPIs, including differences in family practices, language heritage, levels of acculturation to the host society, ostensible socioracial classification, health-related beliefs, and so forth. However, it is not clear to what extent “newcomers” reflect the cultures of the AAPIs who have lived in the US since before restrictive immigration laws were relaxed.

Much research has been conducted to delineate the cultural parameters of AAPI cultures. One difficulty, however, is that in the quest to obtain large enough sample sizes to pass through the gates of “legitimate” research, researchers either have collapsed across ethnic groups without regard to differences in cultural/racial characteristics or they have generalized results from one of the larger native AAPI groups (e.g., Chinese Americans) to all AAPIs. Either of these strategies is problematic for garnering evidence of the effectiveness of culturally responsive interventions because it is not clear what outcome goals (i.e., evidence) are relevant in

each population and which ethnic groups should or should not be combined to obtain such evidence.

What defines the best available evidence?

The Task Force did not actually define criteria for determining what constitutes “best” or even good evidence or how cultural or racial factors should be incorporated in such evidence. Instead, it listed nine research designs by which such evidence might be accrued (Stuart & Lilienfeld, 2007). To its credit, the Task Force did include four designs that are appropriate for small samples if they are so used (Task Force, 2006, p. 274). These potential small sample designs are (a) clinical observation, which could include case studies; (b) qualitative research, which could be used to describe the psychotherapy experiences of people somewhat from their perspectives; (c) aggregated systematic case studies; and (d) single-case experimental designs. The other designs listed require large samples for acquired evidence to be analyzed properly. For the most part, psychology researchers in the US have relied on large-sample methodologies to acquire evidence of treatment efficacy, although researchers in other health disciplines and countries have been somewhat more receptive to small-sample and mixed-method designs (Apesoa-Varano & Hinton, 2013; Hackethal et al., 2013). Yet virtually none of the research approaches have actually incorporated cultural and racial factors into research design(s), evaluations of practices, and outcome measures except as demographic categories. Moreover, with respect to race specifically, Mir et al. (2012) found that although clinicians and researchers agreed that context should be a focus of interventions involving ethnic minorities, they reached little consensus about whether addressing the effects of racism was necessary.

One example of the obliviousness to racial/cultural factors in studies of evidence-based practice in psychology is apparent in Berke, Rozell, Hogan, Norcross, and Karpiak’s (2011) survey of members of Division 12: The Society of Clinical Psychologists of the American Psychological Association. Berke and colleagues investigated psychologists’ familiarity with online resources for obtaining information about EBP/EST (e.g., Division 12’s empirically supported therapies, Translating Research into Practice [TRIP], National Guideline Clearinghouse). They also studied the psychologists’ understanding and use of research methods by which evidence was presumably accumulated (e.g., “test reliability” [sic], randomized clinical trial designs, structural equation modeling). Of the clinical psychologists ($N = 549$) surveyed, 86.2% were “Caucasian” (by which the authors presumably meant “White” rather than White and Asian).

Clinical psychologists reported most familiarity with “test reliability,” which is a bit alarming given that it suggests clinical researchers do not understand that reliability is a characteristic of a sample’s responses rather than a test itself (Green, Chen, Helms, & Henze, 2011). Furthermore, Berke et al. (2011) did not inquire about resources as they pertained to ALANA/RIG populations specifically. Given the exclusion of this topic from the study by Berke and colleagues and other

studies, it seems unlikely that characteristics of ALANAs and RIGs have been factored into the development of standard measures of symptom manifestations commonly used in outcome research. It is also concerning that measures used to obtain evidence of treatment effectiveness have not been derived from thoughtful studies of the effects of culture and racial socialization on psychometric data and that such topics have not been a focus of traditional psychometric education (Aiken et al., 1990).

Berke et al. (2011) did not explicitly attend to awareness of racial/cultural influences as an element of the psychologists' knowledge base with respect to research methods and statistics. Yet they found that "racial minorities" of unspecified socio-racial or ethnic classifications ($N=76$) in their sample knew more about SAMHSA's (Substance Abuse and Mental Health Services Administration of the federal government) National Registry of Evidence Based Programs and Practice and PsycINFO (the database of the American Psychological Association) than their White counterparts, a finding they attributed to the fact that the former were more recent degree recipients. If these two online resources are more likely than the others studied by the authors to include information pertinent to ALANAs/RIGs, then the researchers who are least likely to be funded are most likely to have relevant knowledge about the racial and cultural life experiences of the populations of interest (Ginther, 2011).

Perhaps the limited or noninclusion of ALANA/RIG researchers in funded research accounts for the failure of traditional clinical researchers and/or psychometricians to recognize the need for new paradigms for evaluating outcomes (i.e., providing evidence) for interventions in these neglected populations. Consequently, gatekeepers hinder the collection of evidence pertaining to culturally responsive interventions by requiring conformance to research standards that might not be appropriate for ALANA/RIG samples (Helms & Henze, et al., 2005). Missing from the discussions on research standards is the consideration of research designs, scale or survey development, and psychometric and statistical analyses that might be appropriate for numerically small ALANA/RIG samples. Also absent are strategies for conceptualizing the effects of racial and cultural socialization on participants' responses to diagnostic, process, and outcome measures.

What is evidence for evaluating effectiveness of culturally responsive intervention?

The same types of criteria that are used to define culturally responsive interventions (CRI) ought to be used to obtain and evaluate evidence of their effectiveness or lack thereof. Bernal, Jiménez-Chafey, and Rodríguez (2009) contend that CRI should focus on the treatments, providers, and the clients or patients. Roysicar (2009) added the additional dimension of relationship quality or alliance. Moreover, whether the treatment is a modification of standard practice or an alternative model for a specific population (Hulme, 2011), some framework is

needed for specifying the type of evidence necessary for evaluating treatment effectiveness.

Examination of evidence of the effectiveness of each of the existing CRI models is beyond the scope of the present paper, but the three meta-analyses that have been conducted on CRIs can be inspected for what they reveal about the strengths and limitations of evidence in CRI (Benish, Quintana, & Wampold, 2011; Griner & Smith, 2006; Huey & Polo, 2008). Of particular interest is discovering how CRI researchers managed issues of research design (e.g., sample size, ethnic diversity) as they interacted with measurement or acquisition of evidence, as well as how racial theoretical constructs and cultural factors were included in their collection of evidence. In meta-analyses, effectiveness of treatment typically is reported as standardized mean differences or, alternatively, standardized differences between means of comparison groups (“ d ”) or numbers of standard deviations separating the groups rather than significance levels. Cohen (1988) defined standard deviations of .2 as small, .5 as medium, and .8 as large.

According to the recommendations of culturally responsive theorists (Bernal et al., 2009; Hulme, 2011; Roysicar, 2009), the collection of adequate evidence should attend to every aspect of the helping process including (a) characteristics of service providers (e.g., therapists), (b) nature of treatment(s), (c) characteristics of clients, (d) the dynamics of the process (e.g., alliance, relationship, or engagement), and (e) outcome. Since meta-analyses often do not offer sufficient information to make fine-grained analyses, when additional details were needed to address the focal issues, one original study was selected for further inspection from references cited by the meta-analysts.

Review of meta-analyses on CRIs

Benish et al. (2011)

The authors compared “culturally adapted psychotherapy” (CAP) to “bona fide psychotherapy” (BFP) in 21 published and unpublished studies (2011, p. 284, Table 1) and to control groups (CG) in 59 studies that had included CAP/CG comparisons. CAP was defined as interventions that were modified to incorporate the client’s cultural beliefs about illness, which they identified as “myth” in the manner of Frank and Frank (1993); BFP was defined as unmodified standard treatments for which there was some evidence of efficacy; and CG was defined as any unadapted comparison treatment regardless of whether it was a bona fide treatment. In general, they found that CAP was more effective than control or comparison treatments ($d = .41$ for primary measures; $d = .33$ for all measures) and that CAP interventions were more effective than nonadapted interventions (i.e., BFP) in the same studies ($d = .32$ for primary measures). The authors defined “primary measures” as those intended to assess specific symptoms (e.g., depression) and “secondary measures” included whatever other measures the author(s) considered appropriate. Banks, Hogue, Timberlake, and Liddle’s (1996) study was

selected from Benish et al.'s Table 1 to obtain supplemental information about the effects of cultural adaptations on evidence.

Characteristics of service providers. No direct information was provided about service providers' (e.g., therapists) cultural characteristics by the meta-analysts. In their analyses of moderators of CAP, Benish et al. (2011) indicated that some studies matched clients and therapists by ethnicity (probably meaning ethnic or racial group), but they did not report how many of these studies there were or what ethnicities or ethnic groups were represented. Also, Banks et al. (1996) did not describe cultural characteristics of their service providers, but did report their demographic characteristics (i.e., socioracial classifications, gender, and education and training).

Nature of treatments. Benish et al. (2011) selected studies if the culturally adapted treatment focused on one or more of six illness conceptual explanations subsumed in the Barts Explanatory Model Inventory (BEMI; Rudell, Bhui, & Priebe, 2009), such as beliefs about the nature of symptoms (somatic, mental, or behavioral), origins of illness, and appropriate treatments. Thus, by inference, each of the analyzed studies must have made some attempt to adapt treatments to conform to clients' cultural health beliefs. However, no information was provided about the fidelity of treatment, that is, how it was determined that the treatment was actually administered as intended in any of the treatment conditions.

Banks et al. (1996) implemented two different social skills training curriculums, one of which included "culturally relevant elements" alone and another which included culturally relevant elements plus an emphasis on Afrocentric values. The authors monitored the integrity of the treatments through ongoing training for small group facilitators, periodic checks of sessions, and monitoring of audiotapes of sessions. The authors developed a treatment adherence scale, which trained raters used to evaluate conformance of facilitators to the treatment protocol. Banks et al. reported that adherence was quite good, ranging from 88% to 92%, although the manner in which the data were analyzed is not reported. Nevertheless, it is evidence that the treatments were implemented as intended.

Characteristics of clients. In the meta-analysis, clients were described primarily in terms of sociodemographic attributes (e.g., gender, racial/ethnic categories). Accordingly, 46% were girls and women, and participants self-identified as African American (21.6%), Asian American (14.3%), "non-Puerto Rican Latino/Hispanic" (33%), and "Puerto Rican Latino/Hispanic" (26.7%). No information was provided about pretherapy or preintervention assessment of clients' cultural beliefs, values, or explanatory models of health. Banks et al. (1996) also described the sociodemographic attributes of their service recipients, African American children (33 girls; 31 boys). Yet their study differed from other studies in the meta-analysis in that they did conduct pretraining and posttraining evaluations of their clients' Afro-centric beliefs using the same measures, although only one of the

measures had a cultural focus. Neither the meta-analysis nor Bank et al.'s study investigated effects of race-related constructs.

Dynamics of process. Analysis of what aspects of the treatment worked or did not work in Benish et al.'s (2011) meta-analysis was accomplished by means of moderator analyses. They found significant effects ($d = .21$) for modifying interventions to address cultural myths (i.e., what clients believe about illness/health), but not for other factors such as language matching, racial/ethnic matching, and severity of disorder. Banks et al. (1996) did not provide any information about the aspects of the process that were therapeutic or not therapeutic, although it might have been possible for them to analyze their audiotapes for this purpose.

Outcome. In the meta-analysis, primary measures were analyzed separately and in combination with all other measures incorporated in each study. Multiple measures within a study were aggregated. The meta-analysts did not specify what primary outcome measures were used, but, in their discussion, they suggest that the sample consisted of outpatients who were treated for anxiety disorders, psychotic disorders, mood disorders, and behavioral disorders. Their definition of primary measures suggests that the role of racial or cultural factors on outcomes was disregarded. Also, they reported that, overall, most participants (86%) completed culturally adapted as well as nonadapted treatments (88%). Attrition or treatment completion is a type of outcome measure (i.e., evidence), but it is not clear how or whether it was used in evaluating effectiveness of outcomes in the meta-analysis.

Banks et al. (1996) used three self-report outcome measures, one of which assessed Afro-centric values, a focus of their intervention. Also, they used two standard measures of symptoms addressed by their intervention, which had been developed on predominantly White samples, but reportedly had been used with some evidence of reliability and validity in samples with diverse backgrounds.

Research design. The average sample size for Benish et al.'s (2011) most rigorous analysis (i.e., CAP vs. BFP) was about 11 participants per condition (range: 6 to 54), a number that is within range of the median of 12 per condition that Kazdin and Bass (1989; as cited in Huey & Polo, 2008) found for clinical trials in their review of literature. On the face of it, small sample trials make research on ALANA/RIG samples more feasible because by definition ALANA/RIGs are small populations and, therefore, yield small samples. Although individual studies of CRI were not examined to determine whether they had been funded by external funding agencies, personal experience indicates that reviewers often reject such small-sample studies involving ALANA/RIGs because of concerns about statistical issues (power), meaningfulness, and generalizability to other populations. Sample sizes in Banks et al.'s (1996) study were 31 for the treatment group (i.e., culturally adapted treatment) and 33 for the comparison group (i.e., nonculturally supplemented treatment), which is reasonable for a between-group comparison using small-sample statistics.

However, potential interactions between cultural/racial factors and Banks et al.'s (1996) measures or indicators of evidence were virtually ignored. Based on their internal-consistency reliability analyses (i.e., Cronbach's alpha; CA) of their sample's responses, the authors modified most of their outcome scales and subscales, including their only measure of culture-specific outcomes. Such an approach is consistent with experts' recommendations (Wilkerson & the APA Task Force on Statistical Inference, 1999), but ignores the fact that, as one of the statistical analyses contained in the family of general linear models (GLM), CA analysis is a large-sample statistic. As such, it is supposed to satisfy the same statistical assumptions as other multivariate GLM analyses (e.g., large samples, normality of responses; Green et al., 2011). When samples are too small (as was the case with Banks et al.'s sample), dropping items from scales to improve CA means that the researcher is imposing a structure on item responses of a specific sample that might not generalize to other samples. In this case, the deleted items might have reflected the "culture" of the group rather than measurement error. If such was the case, then outcome evidence was potentially distorted and subject to misinterpretation.

Griner and Smith's meta-analysis (2006)

As seemingly the first quantitative meta-analysis of culturally responsive interventions, Griner and Smith's (2006) study has been critiqued by several other scholars (Benish et al., 2011; Bernal et al., 2009), but not with respect to the effects of racial/cultural dynamics on the nature of evidence. Therefore, the same quality issues I addressed with respect to Benish et al.'s (2011) meta-analysis merit consideration. Acosta, Yamamoto, Evans, and Skilbeck's (1983) study was selected for purposes of elaboration.

Characteristics of service providers. Only 47% of the studies included in their meta-analysis of 76 studies reported service providers' ethnic and/or socioracial classifications, which included African Americans (34%), Hispanic/Latino/a Americans (29%), Asian Americans (19%), White Americans (10%), and Native Americans (8%). No information was provided about the service providers' cultural or racial attributes or experiences, although, by inference, some unknown percentage must have spoken a language other than or in addition to English. Acosta et al. (1983) did not describe any therapist characteristics, perhaps because the actual cultural adaptation was a videotape rather than a direct interaction.

Nature of treatments. Griner and Smith (2006) indicated that, for most of the cited studies (84%), researchers reported included cultural values or concepts in their adaptations of standard mental health interventions. There is some ambiguity about whether adaptations should be interpreted as inclusion of the clients' values in the treatment or modification of the clients' values to conform to the standard treatment, the latter of which Acosta et al.'s (1983) study seemingly entailed. In the meta-analysis, 50% of the interventions used from two to four

types of interventions and 43% reported using at least five types of interventions. A variety of cultural modifications were reported, which included language (74%) and ethnic (61%) matching; treatment in a facility with a cultural focus (41%); collaboration with cultural informants (38%); and so forth. Acosta et al.'s cultural adaptation was a pretherapy intervention intended to teach clients what to expect in therapy.

Characteristics of clients. Most of the meta-analysis clients were described according to gender (93%), predominantly women, and ethnic/socioracial group (100%). Across studies, the breakdown was African Americans (31%), Hispanic/Latino/a Americans (31%), Asian Americans (19%), Native Americans (11%), White Americans (5%), and nonspecific (3%). No information was presented about clients' preexisting cultural values, mental health beliefs, or racial attitudes or life experiences. Acosta et al. also described their sample according to gender and ethnic/socioracial group, as well as a variety of other sociodemographic characteristics (e.g., social class, marital and employment status). Although the authors indicated that the cultural intervention was "based upon our patients' ethnic and cultural characteristics and needs" (1983, p. 874), the intervention seemed more suited to meeting the needs of the therapists to have patients understand the therapy process. However, Acosta et al. did assess clients' knowledge and attitudes about therapy before the clients participated in the culturally adapted intervention.

Dynamics of process. Examination of possible moderators is the only information about the therapy process that can be gleaned from Griner and Smith's (2006) meta-analysis. Not reporting that therapists and clients were matched by race ($d = .58$) was related to stronger effects than occurred in studies in which matching was attempted, but not mandatory ($d = .31$); studies in which therapists spoke the same non-English language as clients ($d = .49$) yielded stronger effects than studies in which no information about language matching was reported ($d = .21$). These results suggest that purposeful attending to proxies for culture (e.g., language) may be beneficial to clients who speak non-English languages, but somewhat attending to racial categories may not. Acosta et al. described the types of therapy goals that their intervention was intended to accomplish (e.g., "[encourage clients] to be more open, direct, assertive, and self-disclosing with the therapist"; 1983, p. 874), as well as how they were accomplished (cartoon slides and role-played vignettes). Thus, their intervention permitted the possibility of assessment of which aspects of their intervention were most effective, although they did not investigate this question.

Outcome. Griner and Smith (2006) examined six different types of outcome measures (mental health symptoms, substance use/abuse, treatment duration or client retention, social support or positive social behavior, client satisfaction, and combination(s) of the foregoing). One cannot tell from their tables or descriptions whether these types of measures represented different methods of data collection (e.g., self-report, observer evaluation) or merely different kinds of topics addressed

by means of the same research methodology. By whatever means the information was garnered, client satisfaction or evaluation of services yielded the strongest overall effect on outcome ($d = .93$) in Griner and Smith's meta-analysis, but one cannot tell whether their satisfaction ratings were related to the other outcome measures. Self-reports of knowledge about and attitudes toward psychotherapy were the outcome measures used by Acosta et al. (1983), who found that participants in the experimental condition acquired greater knowledge and more positive attitudes toward receiving therapy than participants in the control group. How Griner and Smith classified the outcomes in Acosta et al.'s study could not be determined.

Research design. From the meta-analysis, it is not possible to discern how many people were actually represented in any of the individual analyses, or, for that matter, what racial or cultural aspects of people, whether clients or service providers, were examined in the studies. Perhaps most important is the lack of information about the psychometric properties of any of the assessments used in the studies. For example, at least some authors in the original studies that Griner and Smith (2006) analyzed apparently provided clients' diagnoses, but did not explain how they were obtained or validated; nor was evidence of reliability and validity of culture-focused or adapted standard measures provided for ALANA/RIG samples.

Huey and Polo (2008)

Huey and Polo's meta-analysis was intended to discover whether standard treatments used for samples of mostly "minority" youths (i.e., ALANAs/RIGs) could be classified as *well-established*, *probably efficacious*, or *possibly efficacious* according to Chambless et al. (1996) and Chambless and Hollon's (1998) evidence-based practice criteria and some of Nathan and Gorman's (2008, as cited in Huey & Polo) criteria for evaluating the methodological rigor of research. Huey and Polo's meta-analysis differs from the other two in three respects. First, their study was not designed for the specific purpose of discovering whether culturally adapted interventions worked better than nonadapted treatments. Second, the authors listed measures actually used in the analyzed studies. Finally, they were the only authors to report that culturally adapted treatments did not work better than standard treatments for ALANA/RIG youths.

For the purposes of this analysis, I focused on the section of the authors' meta-analysis that deliberately addressed cultural adaptations of treatments. Huey and Polo's "conservative" definition of a culturally responsive intervention was one that "identified intervention or clinician characteristics that made treatment more appropriate for ethnic minority participants" whereas their "liberal" definition encompassed treatments that appeared, based on supplementary sources (e.g., manuals, book chapters), to have been modified. Neither of the two experimental studies they identified as proving that culturally responsive treatment did not result

in additional benefits to clients beyond what could be accomplished with treatment as usual could be located in time for this analysis (Genshaft & Hirt, 1979; Szapocznik et al., 1986). So, an alternative from their Table 7 (Fantuzzo, Manz, Atkins, & Meyers, 2005), evidence-based treatments with culture-based elements, was selected (pp. 287–288). Fantuzzo et al.'s (2005) study actually did find that their adapted treatment worked better than the control condition, but it must have been a culturally responsive intervention according to the liberal definition since the authors did not mention that it had been altered to incorporate cultural elements in their procedures.

Characteristics of service providers. Huey and Polo (2008) did not describe any characteristics of the service providers, nor did Fantuzzo et al. (2005).

Nature of treatments. In Table 7, Huey and Polo summarized the cultural variations in the delivery of evidence-based treatments in their review. These included aspects such as the race/ethnic classifications of the service providers, the adaptation of treatment manuals to be culturally appropriate, and sensitivity training for therapists. Fantuzzo et al. (2005) described their treatment conditions in some detail, but, as previously noted, did not directly address cultural adaptations. Huey and Polo did not provide information about treatment fidelity. Fantuzzo et al. indicated that fidelity checks were conducted (average 90% adherence), but did not provide information that could be used by an outside observer (e.g., meta-analyst) as defining evidence of what constituted treatment.

Characteristics of clients. Huey and Polo (2008) described the samples used in their meta-analysis (including the culturally responsive intervention studies) according to numbers of clients, socioracial/ethnic group, average age, gender, and pretreatment symptoms or conditions for which treatment was provided (pp. 266–274, Table 3). The same characteristics were described in Fantuzzo et al.'s (2005) study; 100% of their participants were African American (of unspecified ethnicity). They also provided information about the environmental context of children similar to their participants and in which the experimental and control conditions occurred (i.e., Head Start classrooms). Neither set of authors discussed cultural or racial socialization aspects of the studied samples such as the nature of the racial climate in their contexts.

Dynamics of process. The researchers deemed to have engaged in cultural adaptations of standard treatments did not seem to have provided very informative descriptions of the treatment process. Thus, Huey and Polo (2008) provided descriptions of process such as: "Treatment addressed intergenerational cultural conflict" (p. 288). Many equally vague descriptions seemed to have been quoted directly from original sources. Fantuzzo et al. (2005) mostly aimed to ensure that treatments were carried out appropriately and monitored the sessions to evaluate treatment success in this respect.

Outcome. Although Huey and Polo's (2008) meta-analysis summarized the measures used for assessing the outcomes of the clinical trials, the authors did not discuss the cultural appropriateness of measures for the treated populations and/or psychometric evidence pertaining to cultural/racial aspects of the measures. In 30 studies involving ALANA/RIG youths, 42 different measures were used. Only seven measures or types of data collection (e.g., behavior ratings, Self-Report Delinquency Scale) were used more than once; those used only twice were used by the same authors in different studies. Three measures were used in Fantuzzo et al.'s (2005) study; psychometric information was provided for all of them, but, contrary to good practice, all except one referenced other studies rather than calculating appropriate statistics in the study underway (Thompson & Vascha-Haase, 2000).

Research design. Because they intended to identify levels of effectiveness of standard treatments with ALANA/RIG youths, Huey and Polo specified stringent criteria by which studies were included in their analyses. Some criteria were: at least 12 participants per condition with at least 75% being "ethnic minorities," use of "valid and reliable" measures, and randomized clinical trial methodology with clearly described statistical methods. Fantuzzo's (2005) study met these criteria.

How good is the evidence for evidence?

In sum, examination of these meta-analyses and related studies indicates that the question of what constitutes evidence of efficacy in culturally responsive mental health interventions has been almost totally neglected in every phase of the treatment process from beginning to end. Research designs specified no frameworks for assessing ethnicity or internalized racist experiences. The following four types of cultural equivalence (Lonner, 1985) might be useful for conceptualizing culturally and racially adapted evidence:

1. *Functional equivalence*, defined as the extent to which the same ostensible behaviors (e.g., crying) are interpreted similarly in different cultural or racial groups, occur with equal frequency within these groups, and elicit similar reactions from other members of the groups (e.g., nurturance vs. disdain). From the meta-analyses, Acosta et al.'s (1983) manipulation of service recipients' pretherapy expectations was based on an implicit assumption that traditional therapy was not salient in the recipients' home cultures and, thus, was not functionally equivalent.
2. *Conceptual equivalence*, which is the extent to which different behaviors (e.g., seeking treatment from a spiritual healer rather than a therapist) define the same or analogous constructs between groups. Banks et al.'s (1996) development of a treatment designed to reflect Afro-centric values was probably a manipulation of an aspect of conceptual equivalence.
3. *Linguistic equivalence*, whether the language or dialect used during the process and in evaluations of the process and outcome have been adjusted so that it has meaning to the person(s) being assessed. Matching service recipients and

providers by language was the only measurable cultural adaptation in the meta-analytic studies, even though it was usually reduced to nominal categories. Nominal categories are the lowest form of measurement because they cannot cause or explain behaviors.

4. *(Psycho)metric equivalence*, the extent to which intake, process, and outcome measures assess the same constructs at the same levels across cultural groups. None of the meta-analytic studies directly addressed this form of equivalence with the possible exception of Banks et al.'s (1996) study, which entailed modification of scale scores according to traditional psychometric principles.

According to Helms (1992), failure to consider any of the aforementioned types of equivalence places one at risk of measuring and interpreting meaningless artifacts as if they were meaningful, or in other words, committing a "cultural equivalence fallacy." It is worth considering how this fallacy might be avoided throughout the treatment and measurement process.

Cultural/racial equivalence issues among service providers. As an alternative to the cultural invisibility of service providers in the meta-analytic studies, service providers and researchers should examine their own tendencies to impose incongruent interpretations of culture and race on the service recipient and helping process. The *Guidelines on Multicultural Education, Training, Research, Practice, and Organizational Change for Psychologists* (American Psychological Association [APA], 2002) offers researchers and practitioners advice for examining their own stereotypes and biases because such views may result in harm to help seekers through misdiagnosis, harmful treatment, and unethical use of measures.

Cultural equivalence of the helping process. Many of the reviewed studies treated matching of therapist–client racial categories as a type of intervention, but merely pairing participants based on similar phenotypes or histories is not treatment. Providers and researchers should seek to understand the process, symptoms, and outcomes through the eyes of the presumed treatment beneficiaries. Nicolas, DeSilva, Grey, and Gonzalez-Eastep (2006) provide examples of how this was done with patients of Haitian cultural background.

Cultural equivalence and research and outcome measures. The researcher has the responsibilities of managing the integrity of the research design without doing harm to participants (Helms & Henze et al., 2005), assessing the relevant racial/cultural dynamics of the process, and selecting or developing culturally appropriate measures. Table 1 summarizes some of the ways to avoid committing equivalence fallacies as they pertain to the evidence gathered (e.g., treatment fidelity, outcome). The descriptions following the "Traditional" rows illustrate some ways in which evidence of the effectiveness of culturally responsive interventions is routinely obtained insofar as one can tell from the meta-analyses and individual studies. The argument here is that in continuing to interpret evidence obtained in the

Table 1. Treatment and measurement equivalence assumptions for culturally and racially responsive evidence-based practice.

Type of equivalence assumptions		
Intervention	Use of measures	Example
<i>Functional equivalence</i>		
1. Traditional		
Depression symptoms (e.g., melancholia, sadness) are the same regardless of help seekers' cultural-racial dynamics	Standardized self-report measures, normed on predominantly White samples are interpreted in the same way regardless of race/ethnic cultural dynamics	The same cut scores developed on predominantly White samples are used to determine depression in Asian and Black samples
2. Racially responsive		
Depression symptoms may be interpreted or expressed according to racial stereotypes	Effects of racism (e.g., perceived discrimination) should be intentionally measured	Help seekers may hide or not reveal depression symptoms to avoid being racially stereotyped
3. Culturally responsive		
Nature of symptoms may be dependent on ethnic cultural norms which may be misinterpreted	Ethnic cultural norms may determine whether items describe functional behaviors in help seekers' environments	Help seekers may respond to items according to what is socially desirable in their cultural environment
<i>Conceptual equivalence</i>		
1. Traditional		
The service provider determines the meaning of behaviors	If a measure has been used successfully with other samples, then it measures the same symptoms in the current sample	Depressed people manifest symptoms in the same way regardless of race/ethnicity
2. Racially responsive		
Participants' racial life experiences shape their behavior in treatment including responses to measures	Select measures that allow help seekers to share experiences that may be analogous to depression in their racial group	Relative to Whites, Blacks may express depression as anger and Asians may express depression as withdrawal
3. Culturally responsive		
Participants respond to treatment and measures in ways that are culturally congruent	Develop or interpret measures according to what is culturally normative in the participants' culture	Members of more interdependent cultures (e.g., Asians, Blacks) may show depression through responses to items about relationships rather than moods

(continued)

Table 1. Continued

Type of equivalence assumptions		
Intervention	Use of measures	Example
<i>Metric equivalence</i>		
1. Traditional		
Treated and measured concepts are similarly engaging or disengaging across groups	Measures with evidence of good psychometric properties (e.g., reliability) for some samples pertain to all populations	Scale scores are used as if they provided objective evidence about outcomes regardless of race/ethnicity of responders
2. Racially responsive		
Race-related life experiences may be camouflaged in responses to self-report measures	Even if measures show evidence of similar responding between groups (e.g., reliability, factor invariance), different within-group explanations may exist	Scores on standard measures should be adjusted based on scores on measures of racial experiences (e.g., racial microaggressions)
3. Culturally responsive		
Cultural group experiences may influence style of responding	Response styles (e.g., acquiescence, extreme responses) may characterize responses to scales	Asians and Blacks may use mid-points of scales or endpoints more frequently than Whites depending on the perceived nature of the construct
<i>Linguistic equivalence</i>		
1. Traditional		
Service recipients should be able to express their symptoms in the dominant language	Translators or translated measures might be used if available	Even if based on translated responses, scores have the same meaning as in the dominant group
2. Racially responsive		
Members of the same socio-racial group may speak a variety of languages and dialects regardless of their appearance	Inquire about the person's linguistic background and administer measures in the person's preferred language or dialect regardless of race	People classified as African American (for example) may speak English, French, Spanish, and a variety of other dialects as a first language and, therefore, attribute different meanings to English or translated English words
3. Culturally responsive		
Multiple languages exist; people transmit their cultural beliefs through words	Use standard procedures of back translation to assure that the items have the same meaning to participants from the focal group	Some English words and expressions used in depression measures have no literal equivalent in some non-English dominant cultures (e.g., "feeling blue")

traditional manner one risks perpetuating ineffective and perhaps harmful treatments for nondominant populations through, for example, the underestimation of the seriousness of respondents' symptoms (Choi, 2002). The rows labeled "Racial" or "Cultural" within the types of equivalence in Table 1 provide some examples of how standard measures or evidence could be adapted or interpreted to reflect attention to Lonner's (1985) four types of cross-cultural equivalence in treatments and measurement of them.

Conclusions

The present examination of the current state of evidence in quantitative research on culturally responsive evidence-based practice revealed that existing evidence is of dubious quality, particularly with respect to operationally defining race and culture, and has been guided by no consensual principles for integrating measurement of culture and race into treatment processes or outcomes. Researchers, practitioners, and gatekeepers should focus more directly on developing and using culturally and racially adapted indicators of treatment effectiveness—from beginning to end. Lonner's (1985) principles of cultural equivalence were offered as a preliminary paradigm for specifying the intended focus of cultural adaptations of treatments and corresponding measures. However it comes about, service providers and clinical researchers must explicitly attend to conceptualizing race and culture as psychological constructs from the perspectives of all parties involved in the process. Otherwise, the question of whether standard or culturally adapted treatments are more effective for ALANA/RIG populations will never be answered and, perhaps more importantly, the racial/ethnic disparities in mental health care will continue to go from bad to worse.

Acknowledgements

The author would like to thank Ashley Carey, Kelsey Rennebohm (deceased), and Amanda Reyome for their efforts to evaluate the quality of assessment and outcome measures used in the meta-analyses discussed in this article. A version of this paper was originally prepared for the EBP-CC Conference at the University of Michigan, October 2011, convened by Joseph Gone.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Acosta, F. X., Yamamoto, J., Evans, L. A., & Skilbeck, W. M. (1983). Preparing low-income Hispanic, Black, and White patients for psychotherapy: Evaluation of a new orientation program. *Journal of Clinical Psychology, 39*, 872–877.

- Agyemang, C., Aikins, A. D., & Bhopal, R. (2012). Ethnicity and cardiovascular health research: Pushing the boundaries by including comparison populations in the countries of origin. *Ethnicity & Health, 17*, 579–596.
- Aiken, L. S., West, S. G., Sechrest, L., Reno, R. R., Roediger, H. L. III, Scarr, S., . . . Sherman, S. J. (1990). Graduate training in statistics, methodology, and measurement in psychology. *American Psychologist, 45*, 721–734.
- American Psychological Association. (2002). Guidelines on multicultural education, training, research, practice, and organizational change for psychologists. *American Psychologist, 58*, 377–402.
- American Psychological Association (APA) Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychology, 61*, 271–285.
- Apesoa-Varano, E. C., & Hinton, L. (2013). The promise of mixed-methods for advancing Latino health research. *Journal of Cross Cultural Gerontology, 28*, 267–281.
- Banks, R., Hogue, A., Timberlake, T., & Liddle, H. (1996). An Afrocentric approach to group social skills training with inner-city African American adolescents. *The Journal of Negro Education, 65*, 414–423.
- Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. *Journal of Counseling Psychology, 58*, 279–289.
- Berke, D. M., Rozell, C. A., Hogan, T. P., Norcross, J. C., & Karpiak, C. P. (2011). What clinical psychologists know about evidence-based practice: Familiarity with online resources and research methods. *Journal of Clinical Psychology, 67*, 329–399.
- Bernal, G., Jiménez-Chafey, M. I., & Rodríguez, M. M. (2009). Cultural adaptation of treatments: A resource for considering culture in evidence-based practice. *Professional Psychology: Research and Practice, 40*, 361–368.
- Bhopal, R. (1997). Is research into ethnicity and health racist, unsound, or important science? *British Medical Journal, 314*, 1751–1756.
- Bhopal, R., Vettini, A., Hunt, S., Wiebe, S., Hanna, L., & Amos, A. (2004). Review of prevalence data in, and evaluation of methods for cross cultural adaptation of, UK surveys on tobacco and alcohol in ethnic minority groups. *British Medical Journal, 328*, 76–80.
- Cardemil, E. V. (2008). Commentary: Culturally sensitive treatments: Need for an organizing framework. *Culture & Psychology, 14*, 357–367.
- Carter, R. T., Mazzula, S., Victoria, R., Vazquez, R., & Hall, S. (2013). Initial development of the Race-Based Traumatic Stress Symptom Scale: Assessing the emotional impact of racism. *Psychological Trauma: Theory, Research, Practice, and Policy, 5*, 1–9.
- Castro, F. G., Barrera, M. Jr., & Streiker, L. K. H. (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annual Review of Clinical Psychology, 6*, 213–239.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7–18.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Christoph, P., . . . McCurry, S. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49*, 5–18.
- Choi, H. (2002). Understanding adolescent depression in ethnocultural context. *Advances in Nursing Science, 25*(2), 71–85.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N. J., Erlbaum.
- Dimidjian, S., & Hollon, S. D. (2011). What can be learned when empirically supported treatments fail?. *Cognitive and Behavioral Practice, 18*, 303–305.
- Fantuzzo, J., Manz, P., Atkins, M., & Meyers, R. (2005). Peer-mediated treatment of socially withdrawn maltreated preschool children: Cultivating natural community resources. *Journal of Clinical Child & Adolescent Psychology, 34*, 320–325.
- Frank, J. D., & Frank, J. B. (1993). *Persuasion and healing*. Baltimore, MD: Johns Hopkins University Press.
- Gambrill, E. (2006). Evidence-based practice and policy: Choices ahead. *Research on Social Work Practice, 16*, 338–357.
- Genshaft, J. L., & Hirt, M. (1979). Race effects in modifying cognitive impulsivity through self- instruction and modeling. *Journal of Experimental Child Psychology, 27*, 185–194.
- Ginther, D. (2011). *Science: Black applicants less likely to receive research funds*. Retrieved from http://www.aaas.org/news/releases/2011/0818nih_diversity.shtml?sa_campaign-Internal.
- Green, C. E., Chen, C. E., Helms, J. E., & Henze, K. T. (2011). Recent reliability reporting practices in psychological assessment: Recognizing the people behind the data. *Psychological Assessment, 23*, 656–669.
- Griner, D., & Smith, T. B. (2006). Culturally adapted mental health interventions: A meta-analytic review. *Psychotherapy: Theory, Research, Practice, Training, 43*, 531–548.
- Hackethal, V., Spiegel, S., Lewis-Fernandez, R., Kealey, E., Salerno, A., & Finnerty, M. (2013). Towards a cultural adaptation of family psychoeducation: Findings from three Latino focus groups. *Community Mental Health Journal, 49*, 587–598.
- Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47*, 1083–1101.
- Helms, J. E., & Cook, D. A. (1999). *Using race and culture in counseling and psychotherapy: Theory and process*. Boston, MA: Allyn and Bacon.
- Helms, J. E., Henze, K. T., Satiani, A., & Mascher, J. (2005). Ethical issues when White researchers study ALANA and immigrant people and communities. In J. Trimble, & C. Fisher (Eds.), *Handbook of ethical research with ethno cultural populations and communities* (pp. 299–324). Thousand Oaks, CA: Sage.
- Helms, J. E., Jernigan, M. M., & Mascher, J. (2005). The meaning of race in psychology and how to change it. *American Psychologist, 60*, 27–36.
- Huey, S. J., & Polo, A. J. (2008). Evidence-based psychosocial treatments for ethnic minority youth. *Journal of Clinical Child & Adolescent Psychology, 37*, 262–301.
- Hulme, P. A. (2011). Cultural considerations in evidence-based practice. *Journal of Transcultural Nursing, 21*, 271–280.
- Kent, M. M. (2007). Immigration and America's Black population. *Population Bulletin, 62*(4).
- Kirmayer, L. J. (2012). Cultural competence and evidence-based practice in mental health: Epistemic communities and the politics of pluralism. *Social Science & Medicine, 75*, 249–256.
- La Roche, M., & Christopher, M. (2008). Culture and empirically supported treatments: On the road to a collision? *Culture & Psychology, 14*, 333–357.

- Liu, J. J., Davidson, E., Bhopal, R. S., White, M., Johnson, M. R. D., Netto, G., . . . Sheikh, A. (2012). Adapting health promotion interventions to meet the needs of ethnic minority groups: Mixed method evidence synthesis. *Health Technology Assessment, 16*(44).
- Lonner, W. J. (1985). Issues in testing and assessing in cross-cultural counseling. *The Counseling Psychologist, 13*, 599–614.
- Mir, G., Salway, S., Kai, J., Saffron, K., Bhopal, R., Ellison, G. T. H., . . . Sheikh, A. (2012). Principles for research on ethnicity and health: The Leeds consensus statement. *European Journal of Public Health, 23*, 504–510.
- National Institutes of Health (NIH). (2001). *NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research*. Retrieved from http://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm
- Nicolas, G., DeSilva, A. M., Grey, K. S., & Gonzalez-Eastep, D. (2006). Using a multicultural lens to understand illnesses among Haitians living in America. *Professional Psychology: Research and Practice, 37*, 702–707.
- Office of the Surgeon General. (2001). *Mental health: Culture, race, and ethnicity. A supplement to mental health: A report of the Surgeon General*. Washington, DC: Department of Health and Human Services, U.S. Public Health Service.
- Roysicar, G. (2009). Evidence-based practice and its implications for culturally sensitive treatment. *Journal of Multicultural Counseling and Development, 37*, 66–82.
- Rudell, K., Bhui, K., & Priebe, S. (2009). Concept development and application of a new mixed method assessment of cultural variations in illness perceptions: Barts Explanatory Model Inventory. *Journal of Health Psychology, 14*, 336–347.
- Sheu, H., & Lent, R. W. (2007). Development and initial validation of the Multicultural Counseling and Self-Efficacy Scale-Racial Diversity Form. *Psychotherapy: Theory, Research, Practice, and Training, 44*, 30–45.
- Stuart, R. B., & Lilienfeld, S. O. (2007). The evidence missing from evidence-based practice. *American Psychologist, 62*(6), 615–616.
- Szapocznik, J., Rio, A., Perez-Vidal, A., Kurtines, W., Hervis, O., & Satisteban, D. (1986). Bicultural Effectiveness Training (BET): An experimental test of an intervention modality for families experiencing intergenerational/intercultural conflict. *Hispanic Journal of Behavioral Sciences, 8*, 303–330.
- Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics. *Educational and Psychological Measurement, 60*, 174–195.
- U.S. Census Bureau. (2012). *Table 4: Projections of the population by sex, race, and Hispanic origin for the United States: 2015 to 2060*. U.S. Census Bureau, Population Division.
- Wachtel, P. L. (2010). Beyond “ESTs”: Problematic assumptions in the pursuit of evidence-based practice. *Psychoanalytic Psychology, 27*, 251–272.
- Wilkerson, L., & the APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.
- Yeh, C. J. (2012). Working with Asian American and Pacific Islander clients. In M. E. Gallardo, C. J. Yeh, J. E. Trimble, & T. A. Parham (Eds.), *Culturally adaptive counseling skills: Demonstrations of evidence-based practices* (pp. 139–154). Los Angeles, CA: Sage.

Janet E. Helms, PhD, is the Founding Director of the Institute for the Study and Promotion of Race and Culture, and the Augustus Long Professor of Testing and Assessment, and convener of the annual Diversity Challenge at Boston College. She is a Fellow in Divisions 17 (Counseling Psychology), 35 (Psychology of Women), and 45 (Ethnic Diversity) of the American Psychological Association and is a licensed psychologist. Dr. Helms has received several awards for her research and interventions that attend to racial and ethnic cultural dynamics in therapy and assessment. Her published works focus on racial identity theory and measurement, racial and cultural dynamics in social interactions, and methodological issues and cultural barriers in psychotherapy, testing, and assessment.