

Measuring Research Mentors' Cultural Diversity Awareness for Race/Ethnicity in STEM: Validity Evidence for a New Scale

Angela Byars-Winston^{1*} and Amanda R. Butz²

¹Center for Women's Health Research and Department of Medicine, University of Wisconsin–Madison, Madison, WI 53715; ²Wisconsin Institute for Science Education and Community Engagement, University of Wisconsin–Madison, Madison, WI 53715

ABSTRACT

Research mentors are reticent to address, and sometimes unaware of how, racial or ethnic differences may influence their mentees' research experiences. Increasing research mentors' cultural diversity awareness (CDA) is one step toward improving mentoring effectiveness, particularly with mentees from underrepresented racial/ethnic groups in science, technology, engineering, and mathematics fields. The indicators of CDA for research mentors are not yet known. Thus, we developed a scale to assess CDA related to race/ethnicity (CDA–R/E) in research mentoring relationships informed by multicultural counseling theory and social cognitive theory. The validation process was guided by classical test theory and item response theory and involved qualitative data, cognitive interviews, and an iterative series of item testing with national samples of mentors and mentees. Confirmatory factor analysis evidenced validity for a three-factor mentor scale assessing attitudes, behavior, and confidence, and a two-factor mentee scale assessing attitudes and behavior. The mentee version captures mentees' perception of the relevance of culturally aware mentoring (“Attitudes”) and their perception of the frequency of mentor's culturally aware mentoring behaviors (“Behaviors”). Implications for use of the CDA–R/E scale in practice, such as assessing alignment between mentor and mentee CDA scores, and use in future studies are discussed.

INTRODUCTION

The experiences of individuals from underrepresented racial and ethnic groups (UR) in science, technology, engineering, and mathematics (STEM) fields are often marked by isolation, presumptions of incompetence, racism and sexism, over-visibility for being “different,” and invisibility for their intellect (Ong *et al.*, 2011; Puritty *et al.*, 2017). Scholars investigating academic and career persistence for UR mentees in STEM have identified the importance of general factors like mentors establishing rapport with their mentees and the importance of specific factors like mentors acknowledging the influences of racial and ethnic diversity on their mentoring relationships and their mentees' research training experience (Johnson, 2007; Hurtado *et al.*, 2009; Blake-Beard *et al.*, 2011). Despite this evidence on the salience of cultural diversity in STEM, research mentors may be reticent to acknowledge, and sometimes unaware of, how racial or ethnic differences can influence their mentees' research experiences (Prunuske *et al.*, 2013; Butz *et al.*, 2018; Byars-Winston *et al.*, 2019). Mentors who are unaware of or inattentive to cultural diversity factors in their research mentoring relationships may also be unaware of culturally based conflicts (e.g., misaligned expectations) that can compromise effective research training experiences for UR mentees.

The National Academies of Sciences, Engineering, and Medicine (NASEM) concluded that challenges arising from the effects of cultural diversity factors operating in personal interactions, including race/ethnicity, can compromise UR mentees' persistence in STEM pathways (National Research Council, 2011). These challenges can

Derek Braun, *Monitoring Editor*

Submitted Jul 3, 2019; Revised Jan 4, 2021;
Accepted Jan 12, 2021

CBE Life Sci Educ June 1, 2021 20:ar15

DOI:10.1187/cbe.19-06-0127

*Address correspondence to: Angela Byars-Winston
(ambyars@wisc.edu).

© 2021 A. Byars-Winston. CBE—Life Sciences Education © 2021 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

include disaffirming research environments that leave UR students feeling excluded, encountering explicit or implicit biases that reflect cultural stereotypes, and even direct experiences of racism and discrimination within their training programs (Acosta and Ackerman-Barger, 2017; Colón-Ramos and Quiñones-Hinojosa, 2016; Puritty *et al.*, 2017). As such, the NASEM consensus study, *The Science of Effective Mentorship in STEM* (NASEM, 2019), advanced the importance of culturally responsive mentoring wherein mentors value their mentees' cultural identities as well as their science identities.

Scholars have documented that many STEM faculty feel ill-equipped to address cultural diversity in their mentoring relationships, especially with respect to race and ethnicity (Byars-Winston *et al.*, 2019), and are differentially motivated to do so (Butz *et al.*, 2018). Moreover, some STEM faculty are prone to adopt a “color-blind” stance in their mentoring, opting not to address cultural factors at all (Prunuske *et al.*, 2013). However, ignoring cultural diversity dynamics is not an effective strategy for reducing challenges that can come from cultural diversity factors in interpersonal interactions and, in fact, can inadvertently further erode interracial interactions (Holoien and Shelton, 2012). Research mentoring relationships are one context in which the cultural diversity dynamics described herein often emerge in mentoring UR mentees (Butz *et al.* 2018; Byars-Winston *et al.*, 2019) and can contribute to their attrition in STEM (NASEM, 2019). We need to better understand the role of cultural diversity in research mentoring relationships for both mentors and mentees.

The stagnation of UR individuals' participation in STEM over the last four decades continues to be a national concern (Estrada *et al.*, 2016), and the National Institutes of Health (NIH) leadership has called for researchers to identify psychosocial factors that can mitigate barriers to scientific workforce diversity (Valantine and Collins, 2015). We assert that one such factor is cultural diversity awareness (CDA). Specifically, we assert that research mentoring enacted with CDA can increase mentors' acknowledgment of and responsiveness to their mentees' cultural realities and thereby enhance their mentoring effectiveness in support of their mentees' academic and career success. In one study by Haeger and Fresquez (2016), it was found that culturally responsive mentoring had strong, positive correlations with mentees' favorable rating of their mentoring relationships and significant mentee gains, including refined academic and career goals and feeling competent as a researcher. Disciplines like medicine and counseling have articulated CDA in their professions, but we do not yet know what the indicators of CDA are for research mentors. Moreover, we need a standardized measure to test whether or not CDA is an important part of the mentoring relationship. To address these gaps, the purpose of this study was to develop an instrument to assess research mentors' attitudes, behaviors, confidence, and motivation relating to CDA in research mentoring relationships.

CDA refers to an individual's ability to recognize his or her own culturally shaped beliefs, perceptions, and judgments and to be aware of cultural differences and similarities between one's self and others (National Center for Cultural Competence, n.d.). The National Center for Cultural Competence (n.d.) stated that cultural awareness is “the first and foundational element because without it, it is virtually impossible to acquire the attitudes, skills, and knowledge that are essential to cultural

competence.” Burchum (2002) asserted that it is awareness of our own culturally informed beliefs, values, and behaviors that allows us to appreciate how others are shaped by culture and to recognize similarities and differences between one another. Following these assertions, we posit that research mentors' CDA is necessary for them to subsequently enact culturally aware mentoring practices. In this study, we focus on CDA related to the attitudes about, behaviors supporting, and confidence to implement culturally aware mentoring practices in research mentoring relationships with an emphasis on racial/ethnic diversity.

Our development of the CDA measure was informed by existing theory and research findings from the fields of multicultural counseling and teacher education (see Byars-Winston *et al.*, 2018). Much work has been done on related CDA concepts like cultural sensitivity, multicultural awareness, cultural humility, and cultural competence in the training and assessment of mental health and healthcare providers and pre-service teachers (Larke, 1990; Pohan and Aguilar, 2001; Gay, 2002; Prieto, 2012) and undergraduate students (e.g., Wang *et al.*, 2003). We drew on a range of survey instruments tapping variables about acceptance of others who are different from oneself and behaviors in addressing those differences in one's professional practice. We were also informed by tenets of social cognitive theory that assert that individuals are likely to pursue behaviors that they feel confident about and motivated to perform (Bandura, 1997). Building on this theory and research base, we proposed that mentors who are proficient in CDA are sensitive to cultural diversity dynamics when they arise in research mentoring relationships, are willing and motivated to acknowledge them with mentees, and have confidence to do. We also proposed that mentors' CDA proficiency should be measurable by mentees' ratings of mentor CDA, as it is mentees' perceptions of their mentors and their mentoring relationships that matter in their academic and career development (Byars-Winston *et al.*, 2015). Items from existing diversity measures may not translate well to researchers in STEM due to unclear terms (e.g., “cultural competence”; Suarez-Balcazar *et al.*, 2011) or because items are not situated within a research context. Thus, using our two propositions, we developed a CDA measure for research mentors and research mentees, which we outline in the following sections.

DEVELOPMENT OF THE SCALE

The development of this scale consisted of item generation and three phases of pilot testing and item refinement. We describe each of the phases in this section. The processes described here were reviewed and approved by the researchers' institutional review board (protocol no. 2015-1086).

Item Generation

Item content was generated via several processes. In Fall 2014, two members of the research team read publications on the topic of cultural or multicultural competence across several fields (e.g., Pedersen, 1988; Eberly *et al.*, 2007; Fouad *et al.* 2009; Smith, 2013). We reviewed the following measures: Multicultural Awareness-Knowledge Skills Survey (Kim *et al.*, 2003); Multicultural Teaching Competencies Inventory (Prieto, 2012); Cultural Competence Assessment (Suarez-Balcazar *et al.*, 2011); Cultural Awareness measure (Rew *et al.*, 2003); Cultural Diversity Awareness Inventory (Larke, 1990); and

Teacher Dispositions Towards Diversity (Dee and Henkin, 2002). Items from some of these scales were flagged for potential relevance to the CDA scale from which initial topics and/or item stems were adapted or developed. The entire research team then participated in a four-session training on multicultural pedagogy and culturally responsive teaching conducted by an advanced doctoral candidate in curriculum and instruction. Immersion in the K–12 multicultural education research gave the team critical understandings of scholarship that informed our interview protocol and item development.

We interviewed participants regarding their attitudes toward CDA and its role in research mentoring relationships. Interviews were conducted in Spring 2015 with mentors ($n = 25$) and mentees ($n = 33$) who had participated in a summer research experience for undergraduates at a large, midwestern, research-intensive university. Interview transcripts were thematically reviewed by all team members, who then converted themes into questions and statements about CDA and its indicators (for a description of findings, see Byars-Winston *et al.*, 2019). Several themes emerged, including the relevance/irrelevance of cultural diversity, diffusion of responsibility, concern over appearing prejudiced or causing offense, and acknowledgment of the complexity of addressing cultural diversity in research mentoring relationships. Race and ethnicity represented the hardest cultural diversity topic to discuss for most participants. Therefore, the research team decided that the initial scale should measure CDA as it relates to race and ethnicity rather than attempting to capture the many dimensions of cultural diversity in one instrument.

We began to develop the CDA–Race/Ethnicity Version (CDA–R/E) scale with two project consultants, reviewing a collated list of 100+ items for relevance, clarity, and phrasing and further reducing the item list. Next, the research team and the university's survey center evaluated and refined the items, resulting in an initial 33-item survey. Think-aloud surveys were administered to three mentors and three mentees who provided verbal feedback on items as they completed the survey. These participants received a \$20 gift card as compensation for their time.

After the think-aloud stage, it became clear that the scale should be revised to generate a scale more consistent with our four-pronged definition of CDA (attitudes, behaviors, confidence, motivation). The research team also felt that having a similar response scale for each of the four factors and the items contained within those factors would be more useful and easily scored. Using results from structured interviews, internal review of the scale, and think-aloud surveys, the team finalized a list of 49 initial items that were pilot tested.

The CDA–R/E Scale

The CDA–R/E scale for mentors was initially hypothesized to comprise four subscales, each measuring a dimension of CDA as it relates to race/ethnicity: Attitudes (17 items), Behaviors (12 items), Confidence to Enact CDA Behaviors (11 items), and Motivation to Enact CDA Behaviors (9 items).

The Attitudes subscale was designed to capture mentor and mentee attitudes about the place of CDA in the research mentoring relationship. Respondents were asked to think about their research mentoring relationships in general and rate each item on a six-point Likert-type scale of agreement, ranging from

“strongly disagree” (1) to “strongly agree” (6). This subscale was conceived as a stand-alone measure of mentor and mentee CDA attitudes or a way to compare mentor and mentee CDA attitudes.

The Behaviors subscale was designed to assess the extent to which mentors incorporated CDA practices into their research mentoring relationships. Responses were on a six-point Likert-type scale ranging from “never” (1) to “all of the time” (6) regarding how frequently each behavior occurred.

The Confidence subscale was designed to assess the degree of confidence that mentors have in their ability to perform CDA-relevant behaviors. These items were developed based on the concept of self-efficacy, or the belief in one's ability to complete a given task (Bandura, 1997). Mentors were asked to rate their level of confidence on a six-point Likert-type scale from “not at all confident” (1) to “completely confident” (6).

The Motivation subscale was designed to measure mentors' motivation to incorporate CDA-relevant practices into their research mentoring relationships. Mentors rated on a six-point Likert-type scale ranging from “extremely unmotivated” (1) to “extremely motivated” (6) their levels of motivation to incorporate each CDA-relevant behavior into their research mentoring relationships. In addition, mentors were asked to respond to an open-ended prompt that asked them to indicate why they were more or less motivated to do each of the behaviors listed. The Confidence and Motivation subscales were administered to mentors only.

Collection of Validity Evidence

The process of collecting validity evidence for the CDA–R/E scale took place over three phases. In this section, we present validity evidence for the CDA–R/E measure. It was validated with separate mentor and mentee samples through a three-phase process of pilot testing and scale revision. Phase 1 was a small-scale pilot test of items with trainees and mentors. Phase 2 involved piloting the measure with a national sample of trainees and their mentors to examine the psychometric properties of the scale. Phase 3 examined evidence of construct validity for the revised CDA–R/E with a separate sample of trainees and mentors recruited nationally. The scale manual, including all items and psychometric properties, is in the Supplemental Material.

Phase 1: Initial Pilot Test

Participants. Mentees ($N = 113$) were recruited from several sources within a large university in the midwestern United States. An email was sent to any mentee who had participated in a summer research opportunity program between 2011 and 2015, enrolled in a program directed toward undergraduate research scholars, or enrolled in an independent research biology course in the prior 2–3 years. Sixty-one percent of respondents self-identified as female and 28% as male; one individual did not report gender. The majority of mentee participants were White (82%), American Indian/Alaska Native (2%), Asian (18%), African American (5%), Hispanic/ Latino(a) (16%), and other ethnicities (9%) were also represented in this sample. Most mentees completing the survey were in their third or fourth year of undergraduate study (75%).

Mentors ($N = 108$) were identified and recruited via emails sent to participants in summer mentor training workshops at a

large midwestern research university. Over half of respondents (58%) self-identified as female, 39% self-identified as male, and 3% did not report gender. The racial/ethnic makeup of the sample was predominantly White (88%); the remainder identified as American Indian or Alaskan Native (2%), Asian (7%), African American (1%), or Hispanic/Latino(a) (9%), and 13% reported that they belonged to another ethnic group. Thirty-one percent (31%) were graduate students; 16% were postdocs; 24% were tenured faculty members; 7% were nontenured faculty members; 9% were scientists; 11% reported other positions; and 2% did not report their current positions.

Measures. Participants who opted into the study were directed to an online survey that included the CDA items, demographic items, and two validated scales hypothesized to correlate with the CDA-R/E subscales: The Fear of Negative Evaluation Scale (FNE; Leary, 1983) and items from the Scale of Ethnocultural Empathy (SEE; Wang *et al.*, 2003). At the end of each page, participants had the option to include any comments that they wished to make regarding their responses. For examination of evidence of convergent and discriminant validity with the CDA-R/E scale, two validated measures were included in this survey, described next.

The brief version of the FNE (Leary, 1983) assessed the degree to which individuals might be afraid of external negative evaluations (e.g., “I worry about what other people will think of me even when I know it doesn’t make any difference”) as a proxy for social desirability. The brief FNE scale consisted of 12 items, which respondents rated on a five-point Likert-type scale ranging from “not at all characteristic of me” to “extremely characteristic of me” ($\alpha_{\text{mentee}} = 0.931$; $\alpha_{\text{mentor}} = 0.944$). We hypothesized that this scale would have no significant relationship with the Attitudes, Behaviors, or Confidence subscales, but that it would have a positive relationship with the Motivation subscale, thus providing evidence of both discriminant and convergent validity, respectively. This positive relationship was hypothesized because individuals may be motivated to incorporate CDA-R/E practices into mentoring relationships out of fear of appearing prejudiced (Plant and Devine, 1998). Mean total scale scores were calculated. Due to a survey error, the final item of this scale (“I am afraid that people will find fault with me”) was omitted from 33 mentee surveys. In these cases, we calculated the mean based on the completed items.

The SEE (Wang *et al.*, 2003) consists of 31 items assessing four factors: Empathic Feeling and Expression (15 items; e.g., “When I hear people make racist jokes, I tell them I am offended even though they are not referring to my racial or ethnic group”), Empathic Perspective Taking (7 items; e.g., “It is easy for me to understand what it would feel like to be a person of another racial or ethnic background other than my own”), Acceptance of Cultural Differences (5 items, e.g., “I feel irritated when people of different racial or ethnic backgrounds speak their language around me”), and Empathic Awareness (4 items, e.g., “I am aware of how society differentially treats racial or ethnic groups other than my own”). This scale has evidenced acceptable internal consistency as a total scale and within each subscale ($\alpha = 0.91, 0.89, 0.75, 0.73$, and 0.76 , respectively). We examined the items included in the Empathic Feeling and Expression subscale and removed any items that might solicit a socially desirable answer, duplicated other items included in

the subscale, or had relatively lower factor loadings compared with other items in the subscale ($n = 7$), resulting in a final subscale consisting of eight items ($\alpha_{\text{mentee}} = 0.792$; $\alpha_{\text{mentor}} = 0.769$). These reductions helped combat potential survey fatigue for our respondents. These items, in combination with the four items included in the Empathic Awareness subscale ($\alpha_{\text{mentee}} = 0.793$; $\alpha_{\text{mentor}} = 0.762$), served as measures of convergent validity for this study. We hypothesized that both Empathic Feeling and Expression and Empathic Awareness would correlate positively with all four CDA-R/E subscales.

Analyses. We examined item-level data using several criteria to flag problematic items for removal. We examined each item’s mean, SD, and distribution to determine whether responses followed a normal distribution. Using the suggested criteria (West *et al.*, 1995), we flagged items with standardized skewness values greater than an absolute value of 3, or whose standardized kurtosis value was greater than an absolute value of 7. We next flagged any items where all response categories were not used or where items within each subscale were significantly correlated with one another ($p < 0.05$). Subscale internal consistency (i.e., Cronbach’s α) and item-total correlations were used to determine each item’s influence on the scale’s internal consistency. Any item whose removal improved the internal consistency of the scale by more than 0.02, whose removal did not have an impact on the scale’s internal consistency, or whose item total correlation was less than 0.3 was flagged for removal (Kline, 2015). We then examined correlations between each item and the FNE and the two SEE subscales to determine whether they correlated in the expected direction. Any items that did not correlate with at least two of the three scales in the expected direction were flagged for removal. Finally, we examined participants’ overall comfort in completing each subscale and any comments they provided in the survey. This information, in conjunction with the quantitative criteria listed earlier, determined removal or revision of items.

Results. Basic descriptive statistics, internal consistency, and correlations with validated measures expected to correlate with our constructs were examined for both the mentor and mentee scales (see Supplemental Material). Any items that were flagged on at least two item-level criteria were identified for potential removal. These items were discussed by the researchers and were either revised or removed from the subscale. The entire scale was then re-examined to determine whether the remaining items were specific to the mentoring relationship and were theoretically relevant to one another. These initial analyses resulted in removal of 10 items from the Attitudes subscale, four items from the Confidence subscale, and three items from the Motivation subscale.

Based on quantitative and qualitative criteria, the Behavior subscale items needed substantial revision. For instance, some mentee participants stated that behavior items were difficult to answer, because several items required them to infer what their mentor’s cognitive processes might be. The research team discussed culturally aware mentoring practices described in research (Byars-Winston *et al.*, 2018, 2019) and consulted feedback from pilot participants to list behaviors that mentors and mentees alike identified with CDA. Eight new behavior items were generated for the next revised CDA-R/E scale.

The revised scale was given to a group of researchers familiar with theoretical constructs relevant to CDA in mentoring. The group provided feedback on clarity and wording of the items that further refined the scale. This resulted in 28 items measuring CDA-related attitudes (seven items), behaviors (eight items), confidence (seven items), and motivation (six items).

Phase 2. Additional Pilot Testing and Analysis of Psychometric Properties of the Scale

Participants. Mentees ($N = 1070$) were undergraduate and postbaccalaureate students invited to participate in this study via a postconference survey distributed to all attendees at an annual biomedical conference. One-third (33%) of the sample self-identified as male, 66% as female, less than 1% identified with another gender identity, and 1% chose not to report gender. The sample's racial/ethnic makeup was 36% Black or African American, 27% Hispanic or Latino(a), 20% bicultural, 7% Asian, 7% White, and 1% American Indian, Alaska Native, or Native Hawaiian/Pacific Islander; 2% did not report race/ethnicity.

Mentors ($N = 301$) were recruited by email announcements sent to Research Experience for Undergraduate site directors and snowball-sampling techniques. Half reported their gender as male, 49% as female, and 1% reported another gender identity. The racial/ethnic makeup on the sample was predominantly White (84%); 8% identified as Asian, 2% as Black or African American, 5% as Hispanic, 1% as bicultural, and less than 1% as Native American or Alaska Native. Forty percent of mentors indicated that they were tenured faculty members; 23% were nontenured, 16% were graduate students, and 7% were postdoctoral researchers; 7% were lab technicians or scientists, and 7% reported other career stages (e.g., administrators, emeriti, industry professions).

Measures. Mentors completed the revised CDA-R/E that included the Attitudes (seven items), Behaviors (eight items), Confidence (seven items), and Motivation (six items) subscales. Mentees completed the revised CDA that included the Attitudes (six items) subscale. Surveys were administered via Qualtrics, an online survey tool. Mentees had the option to complete the CDA as part of the larger conference evaluation survey.

Analyses. We used classical test theory (CTT) and item response theory (IRT) techniques to examine the scale's psychometric properties. We ran exploratory factor analyses on each subscale using principal axis factoring with varimax rotation. Principal axis factoring is an extraction method that is appropriate for ordinal data that may not be normally distributed (Costello and Osborne, 2005; Knetka *et al.*, 2019). Because we expected each subscale to be unidimensional, we chose varimax rotation to simplify the interpretation of factors in this initial analysis (Field, 2009). Scree plots and eigenvalues were examined to determine how many factors were appropriate to retain. Eigenvalues higher than 1 as well as the inflexion point of the scree plot were considered when determining how many factors to retain for each subscale (Cattell, 1966). These values and visual representations of the factor structure of each subscale were then examined relative to our own assumptions and intentions for the dimensionality of each subscale. Factor loadings were examined, and any items with factor loadings lower

than 0.4 and items that appeared to load on multiple factors were further examined. Because we expected each subscale would produce a unidimensional set of items, any additional factors were further scrutinized to determine whether items represented a truly distinct factor or whether the loadings represented an artificial factor. Descriptive statistics, interitem correlations, item-total correlations, and internal consistency values were examined using the same criteria outlined in phase 1. We also examined response category usage for each item to ensure that all response categories were used.

IRT is a common way to examine psychometric properties of new and existing scales by fitting statistical models to data to examine the relationship between a latent trait (in this case, CDA) and observed responses. Unlike CTT approaches, IRT allows for examination of scale characteristics at the item level (deAyala, 2009; Toland, 2014). Our analyses followed the four-step process outlined by Toland (2014). First, we clarified the purpose of our use of IRT, which was to examine the psychometric properties of the newly developed CDA-R/E scale. Second, we determined which models to examine with each subscale. Given the ordered polytomous nature of the data, we fit the data to a graded-response and a reduced graded-response model (Samejima, 1969). The first model allows for unique slope parameters, whereas the second model constrains this parameter to be equal across all items. Third, we inspected the data. As noted earlier, we examined the number of responses in each response category for each subscale to determine whether any categories were underutilized. Finally, we examined whether the data met all assumptions for IRT analysis, specifically, dimensionality, local independence, and model-data fit. Dimensionality was assessed via the exploratory factor analyses described earlier. The local independence assumption was examined in the IRT model by assessing the local dependency (LD) statistics provided in IRTPRO (LD χ^2 ; Scientific Software International, 2011). The local independence assumption indicates that other items within the scale or other latent traits are not influencing item responses. Items with LD statistics larger than $|10|$ were examined and considered for removal. Model-data fit was assessed using the item fit statistics and option-response function plots provided by IRTPRO. Items with poor model-data fit or whose option-response function plots indicated that response categories were not being used as expected by each model (i.e., categories should appear in the expected order with separation between each category threshold) were removed.

RESULTS

We first examined the Attitudes subscale by combining the mentor and mentee data. An initial exploratory factor analysis revealed that three items were loading on a separate factor. As a result, we removed these items to obtain a unidimensional scale. The graded-response model was determined to best fit the data; however, results revealed poor model-item fit and local dependency statistics that were larger than the threshold of $|10|$ that we had set before analysis. Further examination of the data revealed that the fit and local dependency issues discovered in this analysis were primarily from the mentee data. Therefore, we proceeded to conduct a separate analysis for the Attitudes subscale using only the mentor data. Because we intended to conduct additional pilot testing of the scale with

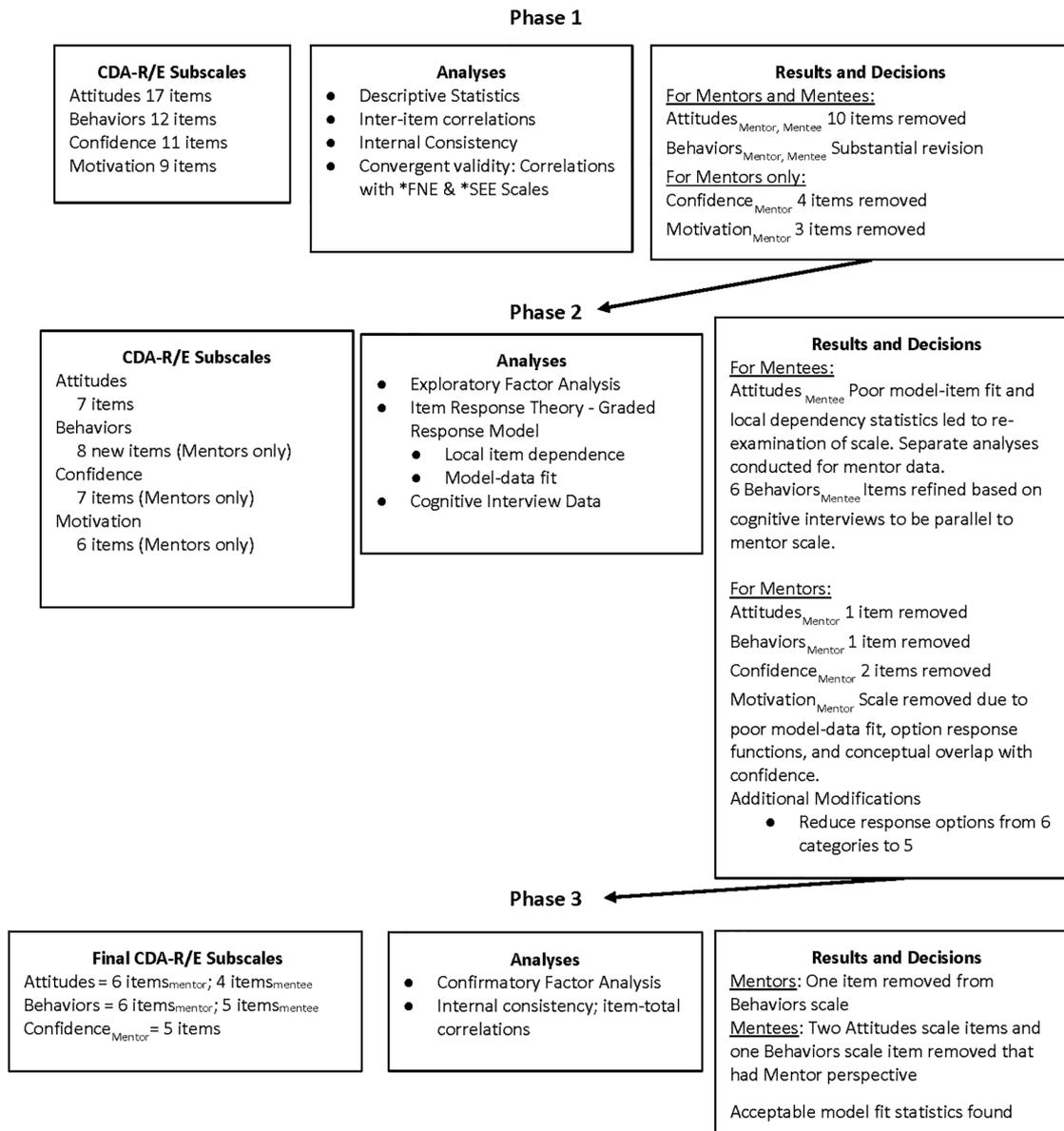


FIGURE 1. Validation phases for the CDA–R/E measure: mentor and mentee versions. FNE, Fear of Negative Evaluation; SEE, Scale of Ethnocultural Empathy (Wang et al., 2003).

another sample of mentors and mentees, we decided to retain the items in the Attitudes subscale pending further analysis.

Mentors

A separate exploratory factor analysis using only the mentor data was run to examine the properties of the scale; with this sample, exploratory factor analyses revealed all subscales to be unidimensional. Overall, the graded-response IRT model was determined to provide the best fit with the data. Items with lower than expected factor loadings, interitem correlations, poor model–data fit, local dependency, or problematic option-response functions were flagged for further analysis.

Results of option-response functions revealed the need to collapse response categories from six to five for the Attitudes, Behaviors, and Confidence subscales. One item (“Race/ethnic-

ity has an impact on the relationship between the mentor and a mentee”) was removed from the Attitudes subscale due to low factor loadings and interitem correlations. One item (“I go outside of my comfort zone to help mentees feel included in the lab”) was removed from the Behaviors subscale due to the ordering of response categories reflected in the option-response function. Two items were removed from the Confidence subscale. The first item (“Identify how the privilege attached to racial/ethnic identities influences the mentoring relationships [e.g., norms, expectations, communication style]”) was removed due to the ordering of response categories reflected in the option-response function. The second item (“Ask questions about a racial/ethnic experience when I do not understand”) was removed due to poor model–data fit. Poor model–data fit statistics and option-response functions for the Motivation

subscale led us to further examine this subscale as a whole. Based on conceptual overlap between motivation and confidence (i.e., self-efficacy; Bandura, 1997) we removed the Motivation subscale before further pilot testing. An overview of the results from phase 2 is presented in the Supplemental Material.

Phase 2b. Think-Aloud Cognitive Interviews

During phase 2b of pilot testing, a complementary CDA-R/E Behaviors scale for mentees was developed based on the items that had been revised for use with mentors. This subscale was conceptualized as a measure to assess the degree of alignment between mentors' self-reports of enacting CDA-R/E behaviors and mentees' perception of mentor's CDA-R/E-promoting behaviors. We conducted a second round of cognitive interviews with a new cohort of mentees in Summer 2016 with the revised CDA-R/E Behaviors scale that paralleled the one previously developed for mentors. We asked mentees to highlight items that were unclear and to debrief their experience with the survey as part of a focus group. Participants ($N = 41$) were undergraduate students enrolled in a summer research experience at a large midwestern research university; 54% identified as female and 46% as male. Fourteen percent identified as Black or African American; 12% as Asian; 32% as White; 2% as American Indian or Alaskan Native; 37% as Hispanic or Latino; and 2% identified with more than one race or ethnicity.

Participants completed the survey and were asked to write down their reactions or comments to items as they completed the survey. Once all participants had completed the survey, they participated in a debriefing session with a researcher on the study team. Notes from the debriefing session and written comments collected on surveys were used to further refine items before the next phase of pilot testing.

Phase 3. Evidence of Construct Validity of the CDA-R/E Scale for Mentors and Mentees

Participants. The final round of pilot testing involved a new cohort of mentees and mentors who completed the revised scale as part of a survey on their mentoring relationships. Mentees ($N = 725$) were undergraduate and postbaccalaureate students invited to participate in this study via a post-conference survey distributed to all attendees at an annual biomedical conference. Sixty-eight percent of participants identified as female and 31% as male; less than 1% identified with another gender identity. Thirty-four percent of participants identified as Hispanic or Latino; 35% as Black or African American; 14% identified with more than one race or ethnicity; 7% as Asian; 7% as White; and less than 1% identified as American Indian/Alaskan Native or Native Hawaiian or Other Pacific Islander; 2% did not report race or ethnicity.

Mentors ($N = 275$) were individuals who completed the CDA-R/E instrument as part of a survey evaluating research experiences for undergraduates and individuals who responded to an invitation to participate via snowball sampling. Sixty percent identified as female; 38% as male; 2% reported another gender identity or did not report gender. Seven percent (7%) of mentors identified as African American; 10% as Asian; 11% as Hispanic; 68% as White; 6% indicated multiple racial/ethnic identities; 9% did not report racial or ethnic identity.

TABLE 1. Descriptive statistics for final version of the CDA-R/E scale

| Subscale | Mentor | | Mentee | |
|------------|--------|--------------|--------|---------------|
| | N | M (SD) | N | M (SD) |
| Attitudes | 275 | 3.821 (.703) | 723 | 3.541 (.846) |
| Behaviors | 270 | 3.374 (.765) | 708 | 3.075 (1.154) |
| Confidence | 270 | 3.767 (.669) | | |

Measures. The revised version of the CDA-R/E scale was administered to mentors and mentees via an online survey tool. Mentors completed the Attitudes (six items), Behaviors (seven items), and Confidence (five items) subscales. Mentees completed the Attitudes (five items) and Behaviors (five items) subscales. There was one less Attitude item administered to mentees due to a survey error; two items on the Behavior subscale were from the mentor's perspective and would thus be difficult for mentees to answer (e.g., "I reflected upon how the research experience might differ for mentees from different racial/ethnic groups"). Mentees had the option to complete the CDA-R/E subscales as part of a larger survey evaluating their experiences at the conference.

Analyses. Descriptive statistics and interitem correlations were calculated, including the mean scores and SDs for both the mentor and mentee samples (see Table 1). Internal consistency statistics were calculated using Cronbach's α coefficient, and any items where the internal consistency of the scale would improve with the removal of the item or the item-total correlations were less than 0.3 were flagged for additional analysis. To confirm the scale's factor structure, we ran a confirmatory factor analysis (CFA) using the weighted least-square mean and variance adjusted (WLSMV) estimator available in *Mplus* (Muthén and Muthén, 2017). WLSMV is preferable to the more traditional maximum-likelihood (ML) estimator due to the categorical nature of Likert-type data. We examined fit statistics using chi-square, root-mean-square error of approximation (RMSEA), and comparative fit and Tucker-Lewis indices (i.e., CFI and TLI) using common criteria for determining goodness of fit; $RMSEA \leq 0.05$; $CFI \geq 0.95$; $TLI \geq 0.95$. Though sample size recommendations for CFA vary widely across the literature, the samples in our study are in line with the recommendations by Moshagen and Musch (2014) that models using the robust WLS estimator (appropriate for Likert-type data) with five response categories and between four and six indicators per factor (minimum 0.50 factor loading) will converge properly 100% of the time with a sample size of at least 200. Similar sample size recommendations are also found for ML estimators (see Wolf *et al.*, 2013).

Results. The factor loadings and internal consistency statistics for the CDA-R/E Scale for Mentors and for Mentees are provided in Table 2.

Evidence of Construct Validity of the CDA-R/E Scale for Mentors. A CFA using the WLSMV estimator in *Mplus* statistical software revealed that a three-factor solution was a good fit with the data. The initial fit statistics for the scale were $\chi^2(132) = 284.172$, $p < 0.001$, $RMSEA = 0.065$, $CFI = 0.967$. This finding was consistent with the hypothesized structure for mentors

TABLE 2. Final factor loadings for the CDA–R/E scale^a

| Item ^b | | Mentor scale factor loading | Mentee scale factor loading |
|---|--|-----------------------------|-----------------------------|
| Attitudes subscale ($\alpha_{\text{mentor}} = 0.857$; $\alpha_{\text{mentee}} = 0.797$) ^c | | | |
| A1 | It is important to consider the mentee's and the mentor's race/ethnicity in mentoring relationships. | 0.756 | 0.780 |
| A2 | Mentoring someone with a different racial/ethnic background benefits the research (e.g., exposure to new ideas). | 0.702 | — |
| A3 | It is important for mentors and mentees to talk together about the mentee's racial/ethnic background. | 0.875 | 0.753 |
| A4 | It is important for mentors and mentees to discuss how race/ethnicity impacts the mentee's research experience. | 0.890 | — |
| A5 | My racial/ethnic identity is relevant to my research mentoring relationships. | 0.747 | 0.792 |
| A6 | Racial/ethnic differences between mentors and mentees enrich the research mentoring relationship. | 0.689 | 0.667 |
| Behaviors subscale ($\alpha_{\text{mentor}} = 0.833$; $\alpha_{\text{mentee}} = 0.877$) ^d | | | |
| B1 | I created opportunities for my mentees to bring up issues of race/ethnicity as they arose. <i>My mentor created opportunities for me to bring up issues of race/ethnicity as they arose.</i> | 0.772 | 0.846 |
| B2 | I encouraged mentees to think about how the research relates to their own lived experience. <i>My mentor encouraged me to think about how the research related to my own lived experience.</i> | 0.634 | 0.705 |
| B3 | I reflected upon how the research experience might differ for mentees from different racial/ethnic groups. | 0.652 | — |
| B3 | <i>My mentor was willing to discuss race and ethnicity, even if it may have been uncomfortable for him/her.</i> | — | 0.873 |
| B4 | I raised the topic of race/ethnicity in my research mentoring relationships when it was relevant. <i>My mentor raised the topic of race/ethnicity in our research mentoring relationship when it was relevant.</i> | 0.864 | 0.894 |
| B5 | I implemented specific strategies to address racial/ethnic diversity in my research mentoring relationships. | 0.813 | — |
| B6 | I approached the topic of race/ethnicity with my mentee(s) in a respectful manner. <i>My mentor approached the topic of race/ethnicity with me in a respectful manner.</i> | 0.522 | 0.766 |
| Confidence subscale ($\alpha_{\text{mentor}} = 0.823$) ^e | | | |
| SE1 | Discuss with mentees how it feels to be a minority in science. | 0.667 | — |
| SE2 | Take advantage of opportunities to address race/ethnicity in the research mentoring relationship. | 0.852 | — |
| SE3 | Recognize aspects of the research experience (e.g., lab, fieldwork) that may make racial/ethnic minority students feel vulnerable to confirming stereotypes. | 0.729 | — |
| SE4 | Provide opportunities for mentees to talk about their racial/ethnic identity as it relates to their research experience should the occasion arise. | 0.796 | — |
| SE5 | Notice interactions in the mentoring relationship that could be insulting or dismissive to mentees because of their race/ethnicity. | 0.692 | — |

^aAll factor loadings were significant at $p < 0.001$.

^bIn cases where mentee items were not parallel to mentor items, the alternate wording is provided in italics.

^cFor Attitudes items, mentors and mentees were asked: "Please indicate how much you disagree or agree with each of the following statements"; responses could range from 1 (strongly disagree) to 5 (strongly agree).

^dFor Behaviors items, mentors were asked: "Please indicate how frequently each of the following has occurred in your research mentoring relationship"; mentees were asked: "Please indicate how frequently each of the following occurred in your relationship with your primary research mentor"; responses could range from 1 (never) to 5 (all the time).

^eFor Confidence items, mentors were asked: "How confident are you in your ability to do the following in your research mentoring relationships?"; responses could range from 1 (not at all confident) to 5 (completely confident).

measuring CDA Attitudes, CDA Behaviors, and CDA Confidence, respectively. Based on an examination of the interitem correlations, one of the Behavior items was removed from the mentor subscale. The final fit statistics for the CDA scale were $\chi^2(116) = 220.296$, $p < 0.001$, RMSEA = 0.057, CFI = 0.976. The internal consistency of the final subscales ranged between $\alpha = 0.823$ and $\alpha = 0.857$. The correlations between each of the subscales were significant and positive, r values = 0.387 to 0.537.

Evidence of Construct Validity of the CDA–R/E Scale for Mentees. Based on interitem correlations and an examination of the internal consistency statistics for each hypothesized subscale, one item was removed from the Attitudes subscale, resulting in a four-item scale for mentees. A CFA using the WLSMV estima-

tor in *Mplus* statistical software revealed that a two-factor solution was a good fit with the data, $\chi^2(26) = 95.71$, $p < 0.001$, RMSEA = 0.06, CFI = 0.991. This finding was consistent with the hypothesized structure of the mentee scale (i.e., a subscale for attitudes and behaviors, respectively). The internal consistency of the Attitudes and Behaviors subscales were $\alpha = 0.797$ and $\alpha = 0.877$, respectively. The correlation between the two subscales was significant and positive, $r = 0.282$, $p < 0.001$.

DISCUSSION

In response to the increased calls to improve the training and mentoring experiences of students in STEM, particularly those students from UR groups (National Research Council, 2011; Estrada *et al.*, 2016; NASEM, 2019), we introduced and

validated a scale to assess CDA related to race/ethnicity (CDA–R/E) for use with research mentors and mentees. Through an iterative series of pilot tests, we analyzed an initial set of 49 items using CTT and IRT techniques. The finalized scales (18 items for mentor scale, nine items for mentee scale) show promise as tools to raise awareness of cultural diversity matters in relationships.

Good intentions are clearly not enough to tackle cultural diversity dynamics in research mentoring relationships. Evidence-based efforts are needed to increase mentors' capacity to attend to cultural diversity in these relationships (Byars-Winston *et al.*, 2018), and several such efforts exist. Rew *et al.* (2014) summarized strategies shown to increase cultural awareness, including cross-cultural immersion experiences (e.g., international study), seminars on cultural awareness, completing culture-focused courses, service-learning projects, and reading and critiquing relevant research findings. Studies have also documented the effectiveness of using computer games and simulations to increase cultural awareness. One computer game called *FairPlay*, in which players assume the role of an African-American male graduate student in STEM, has demonstrated effectiveness in significantly increasing players' awareness of implicit bias in STEM settings and their empathy toward UR students (Gutierrez *et al.*, 2014).

To understand and assess the role of cultural diversity in research mentoring relationships, valid and standardized measures are needed to capture the beliefs of mentors and mentees alike about the related attitudes and behaviors in their mentoring interactions. Addressing this need, we developed the CDA–R/E scales for use in mentor assessment, evaluating the effectiveness of training interventions, and advancing research on STEM mentoring relationships. First, reading the CDA item content in and of itself can be a self-assessment to prompt mentors' reflection on their mentoring practices and spark consideration of new ways that they can acknowledge cultural diversity in their research mentoring relationships (Pfund *et al.*, 2014). Second, researchers may be interested in comparing pre and post CDA–R/E scores in response to mentor and mentee training interventions. Third, the CDA–R/E scale can be used to determine alignment or misalignment between mentor and mentee views of CDA for race/ethnicity. For instance, the CDA–R/E measure could be used during regular mentoring meetings for mentors and mentees to assess and give feedback on their CDA ratings or as a discussion tool only, rather than actually rating each party, given the power differential between mentors and mentees. If differences are identified, the mentor could access mentorship resources and tools to facilitate resolution of the differences, like those included in the NASEM *Science of Effective Mentorship in STEMM* Online Guide (www.nap.edu/resource/25568/interactive). Finally, future research should empirically investigate whether CDA for race/ethnicity is an important part of the mentoring relationship. Research questions may include: Does mentor CDA for race/ethnicity moderate the impact of research mentoring relationships on mentee academic and career outcomes? Are mentee ratings of their mentors' CDA for race/ethnicity associated with mentees' perceptions of their own research-related beliefs or with mentees' ratings of their mentors' effectiveness?

We note several limitations to our methodology. We recruited participants for all three phases of pilot testing using convenience and snowball-sampling techniques. Despite our efforts

to recruit large numbers of mentors, the mentor sample sizes were lower in comparison to the mentee sample sizes. We encourage further validation of the scale with additional populations. Further validation is especially important given that most of our mentor samples self-identified their race as White. Although most research mentors in U.S. academic STEM settings are identified as White, Asian, or non-UR individuals (Gibbs *et al.*, 2016), we see value in the CDA–R/E scales being used across all research training settings, including predominantly White and “minority-serving” institutions. Thus, continued research with a more racially and ethnically diverse pool of mentors is needed to confirm whether this scale operates similarly across cultural groups. Similarly, we believe that further analyses of the CDA–R/E scale with larger mentor and mentee samples using IRT approaches will provide additional information on the psychometric properties of this scale and provide additional validity evidence for response processes. We also acknowledge that there are many forms of individual cultural diversity, including gender, socioeconomic status, physical ability status, and so on, and we encourage development of different versions of this scale to capture other dimensions of cultural diversity beyond race/ethnicity.

Overall, the final scale presents matched items for mentors and mentees regarding CDA attitudes and behaviors that can be used in mentoring dyads to assess alignment between research mentors' and mentees' beliefs and observed behaviors. Although we were not able to examine matched pairs in our data, we hope that researchers and practitioners in STEM will find this feature useful.

ACKNOWLEDGMENTS

We are grateful to Shameka Powell, PhD, Richard McGee, PhD, Nadya Fouad, PhD, Jenna Rogers, PhD, Janet Branchaw, PhD, Christine Pfund, PhD, Nancy Thayer-Hart, MS, and Kimberly Spencer, MS, all of whom provided valuable insight and constructive feedback on this scale at various stages of development. This work was supported by the NIH under grant R01 GM094573. Work reported in this paper was also supported by the NIH Common Fund and Office of Scientific Workforce Diversity under grant U54 GM119023 (National Research Mentoring Network [NRMN]). Additional support was received from the Department of Medicine, University of Wisconsin–Madison (UW). The work is the sole responsibility of the authors and does not necessarily represent the official views of the NIH or UW. Portions of this work were presented at the annual meeting of the Association for Psychological Science, Chicago, IL, May 2016, and the Understanding Interventions Conference, San Antonio, TX, March 2017.

REFERENCES

- Acosta, D., & Ackerman-Barger, K. (2017). Breaking the silence: Time to talk about race and racism. *Academic Medicine*, *92*, 285–288. <https://doi.org/10.1097/ACM.0000000000001416>
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Macmillan.
- Blake-Beard, S., Bayne, M. L., Crosby, F. J., & Muller, C. B. (2011). Matching by race and gender in mentoring relationships: Keeping our eyes on the prize. *Journal of Social Issues*, *67*, 622–643.
- Burchum, J. L. R. (2002). Cultural competence: An evolutionary perspective. *Nursing Forum*, *37*, 5–15.
- Butz, A. R., Spencer, K., Thayer-Hart, N., Cabrera, I. E., & Byars-Winston, A. (2018). Mentors' motivation to address race/ethnicity in research

- mentoring relationships. *Journal of Diversity in Higher Education*, 12(3), 242–254. doi: 10.1037/dhe0000096
- Byars-Winston, A., Branchaw, J., Pfund, C., Leverett, P., & Newton, J. (2015). Culturally diverse undergraduate researchers' academic outcomes and perceptions of their research mentoring relationships. *International Journal of Science Education*, 37, 2533–2554.
- Byars-Winston, A., Leverett, P., Owen, A., Benbow, R., Pfund, C., Branchaw, J., & Thayer-Hart, N. (2019). Race and ethnicity in biology research mentoring relationships. *Journal of Diversity in Higher Education*, 13(3), 240–253. <https://doi.org/10.1037/dhe0000106>
- Byars-Winston, A., Womack, V., Butz, A. R., McGee, R., Quinn, S. C., Utzerath, E., ... & Thomas, S. B. (2018). Pilot study of an intervention to increase cultural awareness in research mentoring: Implications for diversifying the scientific workforce. *Journal of Clinical and Translational Science*, 2, 86–94.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245–276.
- Colón Ramos, D., & Quiñones-Hinojosa, A. (2016). *Racism in the Research Lab*. *New York Times*. Retrieved January 30, 2017, from <http://blogs.nytimes.com/2016/08/04/racism-in-the-research-lab/>
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10, 1–9.
- deAyala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford.
- Dee, J. R., & Henkin, A. B. (2002). Assessing dispositions toward cultural diversity among preservice teachers. *Urban Education*, 37, 22–40.
- Eberly, J. L., Rand, M. K., & Connor, T. O. (2007). Analyzing teachers' dispositions towards diversity: Using adult development theory. *Multicultural Education*, 14, 31–37.
- Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutierrez, C. G., ... & Zavala, M. E. (2016). Improving underrepresented minority student persistence in STEM. *CBE—Life Sciences Education*, 15(3), es5. <https://doi.org/10.1187/cbe.16-01-0038>
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage.
- Fouad, N. A., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M. B., ... & Crossman, R. E. (2009). Competency benchmarks: A model for understanding and measuring competence in professional psychology across training levels. *Training and Education in Professional Psychology*, 3, S5–S26.
- Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, 53, 106–116.
- Gibbs, K. D., Basson, J., Xierali, I. M., & Broniatowski, D. A. (2016). Decoupling of the minority PhD talent pool and assistant professor hiring in medical school basic science departments in the US. *eLife*, 5. doi: 10.7554/eLife.21393
- Gutierrez, B., Kaatz, A., Chu, S., Ramirez, D., Samson-Samuel, C., & Carnes, M. (2014). "FairPlay": A videogame designed to address implicit bias through active perspective taking. *Games for Health Journal*, 3. doi.org/10.1089/g4h.2013.0071
- Haeger, H., & Fresquez, C. (2016). Mentoring for inclusion: The impact of mentoring on undergraduate researchers in the sciences. *CBE—Life Sciences Education*, 15(3), ar36. doi: 10.1187/cbe.16-01-0016
- Holoien, D. S., & Shelton, J. N. (2012). You deplete me: The cognitive costs of colorblindness on ethnic minorities. *Journal of Experimental Social Psychology*, 48, 562–555. <https://doi.org/10.1016/j.jesp.2011.09.010>
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55.
- Hurtado, S., Cabrera, N. L., Lin, M. H., Arellano, L., & Espinosa, L. L. (2009). Diversifying science: Underrepresented student experiences in structured research programs. *Research in Higher Education*, 50, 189–214.
- Johnson, A. (2007). Unintended consequences: How science professors discourage women of color. *Science Education*, 91, 805–821.
- Kim, B. S. K., Cartwright, B. Y., Asay, P. A., & D'Andrea, M. J. (2003). A revision of the Multicultural Awareness, Knowledge, and Skills Survey—Counselor edition. *Measurement and Evaluation in Counseling and Development*, 36, 161–180.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford.
- Knetka, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, 18, 1–17.
- Larke, P. J. (1990). Cultural diversity awareness inventory: Assessing the sensitivity of preservice teachers. *Action in Teacher Education*, 12, 23–30.
- Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation scale. *Personality and Social Psychology Bulletin*, 9, 371–375.
- Moshagen, M., & Musch, J. (2014). Sample size requirements of the robust weighted least squares estimator. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 10, 60–70. doi: 10.1027/1614-2241/a000068
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- National Academies of Sciences, Engineering, and Medicine. (2019). *The science of effective mentorship in STEM*. Washington, DC: National Academies Press. <https://doi.org/10.17226/25568>
- National Center for Cultural Competence. (n.d.). *Cultural awareness*. Retrieved October 18, 2018, from <https://nccc.georgetown.edu/curricula/awareness/index.html>
- National Research Council. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. Washington, DC: National Academies Press.
- Ong, M., Wright, C., Espinosa, L. L., & Orfield, G. (2011). Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics. *Harvard Educational Review*, 81, 172–209.
- Pedersen, P. (1988). *A handbook for developing multicultural awareness*. Alexandria, VA: American Association for Counseling and Development.
- Pfund, C., House, S. C., Asquith, P., Fleming, M. F., Burh, K. A., Burnham, E. L., ... & Sorkness, C. A. (2014). Training mentors of clinical and translational research scholars: A randomized controlled trial. *Academic Medicine*, 89, 774–782.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811–832.
- Pohan, C., & Aguilar, T. (2001). Measuring educators' beliefs about diversity in personal and professional contexts. *American Educational Research Journal*, 38, 159–182.
- Prieto, L. R. (2012). Initial factor analysis and cross-validation of the Multicultural Teaching Competencies Inventory. *Journal of Diversity in Higher Education*, 5, 50–62.
- Prunuske, A. J., Wilson, J., Walls, M., & Clarke, B. (2013). Experiences of mentors training underrepresented undergraduates in the research laboratory. *CBE—Life Sciences Education*, 12, 403–409.
- Puritty, C., Strickland, L. R., Alia, E., Blonder, B., Klein, E., Kohl, M. T., ... & Gerber, L. R. (2017). Without inclusion, diversity initiatives may not be enough. *Science*, 357, 1101–1102.
- Rew, L., Becker, H., Chontichachalalauk, J., & Lee, H. (2014). Cultural diversity among nursing students: Reanalysis of the Cultural Awareness Scale. *Journal of Nursing Education*, 53, 71–76.
- Rew, L., Becker, H., Cookston, J., Khosropour, S., & Martinez, S. (2003). Measuring cultural awareness in nursing students. *Journal of Nursing Education*, 42, 249–258.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34, 1–97.
- Scientific Software International. (2011). *IRTPRO for Windows*. Skokie, IL. <https://issicentral.com/>
- Smith, L. S. (2013). Reaching for cultural competence. *Nursing*, 43, 30–37.
- Suarez-Balcazar, Y., Taylor-Ritzler, T., & Garcia-Ramirez, M. (2011). Development and validation of the cultural competence assessment instrument: A factorial analysis. *Journal of Rehabilitation*, 77, 4–13.
- Toland, M. D. (2014). Practical guide to conducting an Item Response Theory Analysis. *Journal of Early Adolescence*, 34, 120–151.
- Valantine, H. A., & Collins, F. S. (2015). National Institutes of Health addresses the science of diversity. *Proceedings of the National Academy of Sciences USA*, 112, 12240–12242.

- Wang, Y.-W., Davidson, M. M., Yakushko, O. F., Savoy, H. B., Tan, J. A., & Bleier, J. K. (2003). The Scale of Ethnocultural Empathy: Development, validation, and reliability. *Journal of Counseling Psychology, 50*, 221–234.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In Hoyle, R. H. (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*, 913–934. doi: 10.1177/0013164413495237