Full Title: Study Protocol for a Mixed Method, Three-Arm Cluster Randomized Comparative Trial Testing the Impact of Culturally Aware Mentoring Interventions on Research Mentors and Graduate Training Programs

Short Title: Study Protocol for a Mixed Method Study Testing Culturally Aware Mentoring Interventions

Ellyssa Eiring[1,¶], Angela Byars-Winston[1,¶,#a,*], Stephanie C. House[1], You-Geon Lee[1], Richard McGee[2], Remi F. Jones[2], Sylvia Hurtado[3], Amanda Carrasco[3]

[1] Wisconsin Center for Education Research, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

[2] Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States of America

[3] School of Education and Information Studies, University of California, Los Angeles, California, United States of America

[#a] Current Address: Institute for Diversity Science, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

* Corresponding Author

Email: ambyars@wisc.edu

[¶] Both authors (EE and AB) contributed equally as senior first authors.

Abstract

Research mentoring relationships are critical for developing the next generation of scientists. However, research mentors have seldom recognized how cultural differences can influence mentoring relationships, particularly for individuals from backgrounds historically underrepresented in the sciences. To address this need, a mentorship education intervention, the Culturally Aware Mentoring (CAM) workshop, was developed. Preliminary studies showed CAM was effective in promoting culturally aware mentoring principles and behaviors. This paper describes the study protocol for a deeper examination of the impacts of three variations of CAM. We use a mixed method, three arm cluster randomized comparative trial design with 33 sites, paired with in-depth case studies at two institutions. The Theory of Planned Behavior and Organizational Learning provide complementary theoretical frameworks to gauge how mentors and their training environments are impacted by CAM. The primary research questions are: 1) What are the short and longer-term impacts of CAM interventions on research mentors' attitudes and behaviors? 2) What are the broader institutional impacts of CAM on graduate training environments? 3) Why, or why not, do mentors and training environments change? Study sites are biomedical PhD programs at high research universities across the United States. The study targets 600 faculty actively mentoring PhD students and the interventions are delivered virtually and led by trained facilitators. Survey data are collected across the trial and the case studies at four time points: pre-intervention, immediately post-intervention, 6 and 12 months post intervention. Post intervention interviews are conducted with a subsample of participants to examine how and why their mentoring practices changed. The case studies assess organizational changes in those environments following CAM intervention. This study's results will reveal extensive insights into the research questions, laying the foundation for broader deployment of CAM to the biomedical and larger research training community.

Introduction

Research mentoring relationships are a catalyst to unlock the potential of the next generation of scientists. These relationships provide developmental spaces in which students' talents and skills are honed, and their pathways into scientific fields are discovered.[1,2] They are also the contexts in which cultural, social, and psychological factors can frustrate the engagement and persistence of emerging scientists from historically underrepresented (HU) racial and ethnic groups identifying as American Indians/Alaska Natives, Black/African Americans, Hawaiians/Pacific Islanders, and Hispanics/Latinos in the sciences.[3] They report often feeling invisible, unvalued, less competent, targeted for bias, isolated, and marginalized.[4–8] Studies document HU students' encounters with faculty bias in seeking graduate advising, as well as stressors like stereotype threat and lack of faculty awareness about challenges they face.[9,10] However, typical "diversity training" has little impact on bias, and can worsen its impact if implemented without strategies to reduce bias and promote equity and inclusion.[11–14] We assert that culturally responsive mentoring[1] is needed to guide faculty mentors to understand the sources and impact of bias on diverse graduate trainees. Yet, accomplishing this goal is far from easy.

There is a growing realization that there is indeed a science to mentorship and that mentorship is a skill to be taught and learned. The National Academies of Sciences, Engineering, and Medicine (NASEM) consensus study report, *The Science of Effective Mentorship in STEMM,*[1] defined mentorship as a professional, working alliance in which individuals work together over time to support the personal and professional growth, development, and success of the relational partners through the provision of career and

psychosocial support. While the consensus study found that effective and evidence-based curricula for developing core mentorship skills and competencies exist, they fell short of addressing elements of culture and social identities as key factors in mentoring relationships. Beginning in 2014, as part of the National Research Mentoring Network (NRMN) Phase 1 funded by the National Institutes of Health (NIH), our diverse, transdisciplinary team created a novel intervention, called *Culturally Aware Mentoring* (CAM), designed to increase mentors' skills for addressing culture and social identities, especially with mentees from racial, ethnic and social backgrounds different than theirs. The long-term goal of the CAM initiative is to increase access to high quality mentoring relationships for all students, particularly HU students, that will advance their success.

The original CAM intervention consists of two components: 1) a day-long, in-person workshop preceded by engagement with 2) a roughly one-hour self-paced online module called Introduction ("i") to CAM or iCAM.[15] The CAM workshop is led by two facilitators and the original content is about 6 hours, subtracting the time allotted for lunch and breaks. The CAM workshop is embedded with active learning design elements[14] across three sections: (1) The "intrapersonal" section requires participants to reflect on their own racial and ethnic identity. (2) The "interpersonal" section links personal racial and ethnic identity to how cultural diversity between mentors and trainees operates in mentoring relationships. It uses a video about the consequences of being "different", introduces research on bias and stereotyping, and describes implications for being a culturally aware mentor. (3) The "skill building" section introduces eight CAM principles and has participants apply and practice them by role playing a case scenario in which race and ethnicity matters are salient. Participants

conclude training with an action plan, identifying what their first step will be toward applying CAM principles in their mentoring practice.[15]

The iCAM module provides participants with foundational content and context for the CAM workshop, including definitions of key concepts and terms used in the workshop. The module covers four topics: (1) race and privilege, (2) the experiences of scientists from HU groups, (3) the realities of cultural diversity in the sciences, and (4) the role of culturally aware mentoring in trainee outcomes. Each section concludes with self-reflection questions and a section called “Go Deeper” with additional materials should participants wish to explore topics further.

Evaluation data from the pilot study of CAM showed both short and long-term positive impacts of the CAM intervention and the online module has been rated as highly valuable in preparing for the CAM workshop.[15,16] The findings demonstrated the potential of the CAM intervention to substantially change faculty understanding of and willingness to engage in the ‘work’ required to build skills to understand, respect and incorporate cultural differences in their mentoring relationships.[15]

Despite these favorable preliminary data, the CAM pilot study was small in scale, limited by resources in NRMN Phase 1. Moreover, some participants asked if the same results could be reached in a shorter time period, as many department leaders felt too few faculty (the primary audience for CAM) would commit an entire day for a workshop. The CAM team wondered if the same level of impact could be achieved in a half-day version but recognized that wide dissemination was an equally critical goal and, thus, testing the efficacy of a shorter version for possible broader uptake was necessary. With funding in NRMN Phase 2, we are now conducting a rigorous, extensive concurrent

mixed method study to test the comparative outcomes of a shorter CAM workshop versus the original design, placing equal weight on a cluster randomized comparative trial, multiple embedded case studies, and interview based qualitative data.

This paper presents the rationale and methodology for our study of faculty mentoring doctoral students in the biomedical sciences, aimed at optimizing the impact of CAM by investigating the underlying mechanisms of action by which CAM affects faculty and the culture of their graduate training programs and departments. Our primary research questions in the current study are: 1) What are the comparative short and longer-term impacts of CAM interventions on research mentors' attitudes and behaviors? 2) What are the broader institutional impacts of CAM on graduate training departments and programs? That is, how can CAM catalyze organizational change in graduate training over time? 3) Why, or why not, do mentors and training environments change? And what are the underlying mechanisms through which CAM impacts mentors? We hypothesize that the full-day CAM intervention will result in greater changes in research mentors' attitudes, competencies, and mentoring practices than the half-day CAM intervention or online training alone. We also hypothesize that organizational learning and change in graduate departments or programs are more likely to happen under conditions where whole departments of faculty undergo CAM training.

To answer the research questions, we pursue two research aims, employing a site-randomized comparative trial research design with repeated measures and qualitative interviews for Aim I and a multiple-embedded case study for Aim II (see

Figure 1). We use the Theory of Planned Behavior[17] and organizational learning theory[18] to inform our research questions and guide our analyses.

Study Intervention-Testing CAM Efficacy by Dose and Approach

In the present study, we originally proposed comparing three conditions: (1) iCAM plus the full-day CAM workshop; (2) iCAM plus a half-day workshop; and (3) iCAM as a stand-alone module (control condition). We chose not to include a no-treatment control condition because we felt it would be difficult to engage a totally passive participant group and because, in previous years, we had inquiries about using the module as a stand-alone training. Groups randomized into the iCAM alone arm are provided with the option for the full-day intervention 12 months later, after data for the site have been obtained. All sites in the iCAM alone arm, except for one, receive CAM training after the 12-month data collection period.

The half-day workshop includes all three sections in the full day (intrapersonal, interpersonal, skill building), but in an abbreviated format that excludes some activities. The half-day version retained the activities that were rated as most effective by participants in the original full-day version.[15,16] This original design allows us to investigate the comparative outcomes of the full vs. half-day CAM intervention (dose) delivered in-person and the comparative outcomes of facilitated learning delivered in person vs. self-paced learning delivered online (format). However, due to the COVID-19 pandemic, we experienced not only delayed initiation of our study but changes to our design. Specifically, we have adapted all the facilitated content to be delivered in an online format. In the revised study design, we investigate the comparative outcomes of

the full vs. half-day CAM intervention (dose) and the comparative outcomes of facilitated learning vs. self-paced learning (approach).

Given our experience with facilitating mentor training interventions virtually,[19,20] we knew that virtual implementation requires additional time, and content must be spread over multiple sessions to avoid participant fatigue in a virtual learning platform. For this reason, the full-day workshop has been translated into three, 3-hour sessions delivered over three consecutive weeks and the half day workshop is now two, 3-hour sessions delivered over two consecutive weeks (see Table 1). Distributing content in this manner required us to rethink the flow of the workshop, such as how to introduce and conclude each session in engaging ways. The new format affords additional time (i.e., content delivered over 2-3 weeks) and allows for additional intercession reflections. Participants are given specific prompts to consider over the week between sessions, and subsequent sessions begin with a brief discussion of participant reflections to the prompts. These reflection discussions provide a transition back to workshop content. The half-day version was first beta-tested at two institutions and then refined based on participant and facilitator feedback.

In addition to the two primary workshop facilitators-–ideally one racialized as white and one racialized as a person of color–, a third ‘support facilitator’ and someone to provide technical support are added for the online implementation. The technical support person projects and advances slides, manages breakout rooms, and handles other technical needs as they arise. The support facilitator provides an extra set of eyes and ears to monitor the virtual meeting room, communicates with any struggling participants in a breakout or private chat, and provides additional help to the technical

support person. Given that the support facilitator is also trained to implement CAM, they can step in if one of the two primary facilitators are prevented from being present.

Materials and Methods

Study Setting

The study sites are drawn from 33 doctoral universities with Carnegie classification of very high research activity (i.e., R1) across the U.S. with biomedical departments having a minimum of 15% HU PhD student population enrollment. The rationale for requiring graduate training programs to meet the 15% minimum is so that faculty participants would have occasion for immediate use of their cultural awareness skills with HU trainees. Most R1 universities have a predominantly white graduate student population, and thus a high need for diversifying the student body and relevant mentor training. If sites believe they meet study eligibility criteria, the research manager (EE) and researcher (SCH) meets with the site host(s) to share information about the study aims, the site expectations including a signed memorandum of understanding (MOU), and faculty participant expectations. If a host institution remained interested, they confirmed their commitment by sending the research manager their institution-signed MOU and departmental or program demographic information that we use in the randomization process to assign interventions.

Following each set of intervention assignments, a meeting with the site host(s) is scheduled to disclose the site's assigned intervention (i.e., treatment condition) and discuss event logistics and the recruitment process. During the meeting, the study aims and host and participant study expectations are reviewed and potential study intervention dates were discussed. Being the most knowledgeable about the local

faculty academic calendar, site hosts are asked to determine the best dates for the intervention. The study team encourages dates that avoid major academic conferences, religious holidays or observances, and common vacation times. If the site is assigned to a workshop intervention, the researcher (SCH) coordinates with the facilitation team to ensure facilitator availability. Once intervention dates are confirmed, participant recruitment begins.

CAM Facilitator Preparation

All CAM facilitators were part of the NRMN Phase 1 Master Facilitators Initiative (2015-2019), a community of practice in which they co-facilitated mentorship training and met annually for in-person retreats to share best practices.[21] They also provided feedback throughout the CAM development process and most attended a CAM facilitator training in 2017 and implemented earlier versions of the training. Therefore, they have an intimate understanding of the curriculum and continue to build mutual rapport and trust amongst themselves. Significant and sustained facilitator preparation is a critical foundation of the CAM intervention effectiveness.[22]

CAM facilitators received a 1.5 day, in-person facilitator training prior to the start of the current study. This training included practice facilitation, introduction of the new half-day version of CAM being tested in the study, and discussion of trial protocols. A second facilitator training was held online to discuss the new adapted format and to practice delivering the CAM workshop online instead of in person. The CAM study facilitators meet quarterly throughout the study to discuss facilitation issues that arise and to ensure consistency in their implementation. Though nine facilitators were trained,

to minimize variability in implementation, five core facilitators lead all CAM workshops for this study, with the others acting as support facilitators.

Workshop facilitators often have little knowledge of the departmental climate at each site prior to facilitating a workshop. To support their workshop preparedness, the research manager (EE) facilitates discussions with at least two of the assigned three facilitators per site, the site host(s), and the CAM researcher (SCH) regarding campus and departmental climate, participant demographics, and power differentials that may impact confidence and engagement among the workshop participants (e.g., a department chair and faculty member attending together). A standard set of questions guide the discussion, including participants' previous mentor training, participants' motivation to attend, and any recent campus or department incidents that may affect interest or participation. The three workshop facilitators (two primary and one support) receive a summary of de-identified participant demographics and aggregate participant responses to two pre-workshop survey questions: 1) My racial/ethnic identity is relevant to my research mentoring relationships (5 pt scale ranging from strong disagree to strongly agree); and 2) How confident are you in your ability to discuss with mentees how it feels to be a minority in science (5 pt scale ranging from not at all confident to completely confident)?

Participant Eligibility and Recruitment

Participant recruitment begins at least three months prior to the intervention. This time frame allows the study team to answer questions, address concerns, and share pre workshop tasks at a pace that fits with faculty demands and class schedules. Initial communications to all department faculty are sent from the site host, which offers a level

of familiarity given that faculty often receive emails from senders unfamiliar to them. This initial email includes details about the CAM intervention objectives, eligibility qualifications, a hyperlink to the site-specific interest survey, and a PDF attachment of the site-specific interest flier. Each interest flier includes a workshop overview, eligibility qualifications, a sample agenda, and facilitator biographies.

To meet eligibility criteria, faculty mentors are required to 1) be a member of the targeted department or program as determined by the site host; 2) be mentoring at least one graduate student in their research group at the time of enrollment; 3) meet session attendance expectations, and; 4) have participated in 4 or more hours of previous mentor training. Interested faculty without previous mentor training are offered the option to meet this prerequisite by completing the *Optimizing the Practice of Mentoring 101: For Research Mentors of Graduate Students, Fellows, and Early-Career Faculty* online course offered through the University of Minnesota's Clinical and Translational Science Institute.[23] Potential participants complete an online interest survey used as a screening tool prior to registration.

Informed Consent

The University of Wisconsin-Madison Institutional Review Board approved this study. All CAM study participants are required to sign an informed consent that is included in the presurvey. The informed consent form describes the study, includes a list of participation expectations which include survey completion, completing the iCAM online module, and when relevant, participating in the CAM workshop. Also mentioned in the consent was that individuals may be invited to participate in an interview and a description of how participant survey and interview data will be used. In appreciation of

their time, study participants that complete three of the four study surveys receive a $25 electronic gift card and another $25 electronic gift card if they participate in a study interview. Electronic signature on the informed consent confirms participants' registration in the study and in their site's assigned intervention. Combined with securing the MOU described above, we obtain informed consent at both the site (cluster) and individual (participant) levels.

AIM I Randomization

Power analysis for cluster randomized comparative trial - Calculating our sample size

We calculated a sample size with a method that accounts for intraclass correlation, expected effect size, and the power of the study. Based on our pilot study and relevant previous research,[15,24] we expect the effect size in the study to be bigger than 0.35, which is considered small/medium using Cohen's[25] criteria. With three treatment groups, our proposed sample size of 600 mentors from 30 sites (20 mentors x 10 sites for each treatment group) is adequate for the main objective of this study to compare the three CAM treatment groups at 80% power.

Randomization: Minimization and Best Balance

In any trial, it is desirable to have a good balance of prognostic factors among treatment groups. A similar distribution of prognostic factors among treatment groups improves the accuracy and precision of the results, and thus enhances the credibility and acceptance of the results.[26,27] Many randomization techniques are often applied to obtain a good balance of prognostic factors, which prevent selection bias and insure against accidental bias.[28] However, especially in small sample trials or trials with many prognostic factors, substantial imbalances among randomly allocated treatment groups

are likely to occur.[26,27] In fact, cluster randomized control trials often lack sufficient numbers of clusters for many reasons. When the number of clusters is small, many randomization techniques such as simple randomization, matching, and stratification can easily result in unequally distributed cluster characteristics by any chance.[26] Alternatively, covariate adaptive randomization (e.g., minimization, best balance), in which clusters are assigned to the treatment groups in a way of minimizing the imbalance of prognostic factors between treatment groups, has been recommended by many researchers in cluster randomized control trials and shown to improve balance better than other allocation methods.[26,28] We use minimization and best balance randomization techniques in the present study to optimize the balance between treatment groups in terms of prognostic factors.

Rationale: Minimization and Best Balance

We use a process of rolling recruitment to randomly assign all 33 sites (clusters) in our study. Because the best balance method requires all subjects to be known before the randomization process to calculate the imbalance measures from all possible allocations, the best balance alone was not applicable to our study.[26] Thus, we also employ the minimization approach because it provides flexibility to add subsequent subjects (i.e., rolling recruitment) in the randomization process.[29,30] For newly incoming subjects (subsequent subjects), we use minimization to assign them into treatment groups in a way of selecting an allocation with the least imbalance score from all hypothetical allocations of each subject to every group.[31]

To compare the best balance method along with minimization to minimization alone, we simulated two scenarios 500 times for each, assuming 3 treatment groups for

30 simulated sites. For the first scenario, we used the best balance approach for the first 9 confirmed sites and then applied the minimization method to the next 21 sites. For the second scenario, we used only minimization for 30 sites. After creating a simulation dataset (30 sites with 9 randomly generated prognostic factors), we calculated the imbalance scores of 500 simulated allocations for each scenario and compared the mean difference. Test statistics showed that allocations from the first scenario were statistically more balanced than allocations from the second scenario ($p < 0.001$). Our simulation analysis suggested that a new approach, the best balance method within the minimization framework, outperformed the minimization framework alone.

Measure of Imbalance

To optimize the balance between treatment groups, a measurement of imbalance is required. The overall imbalance is often calculated as the sum of the squared imbalance for each category of each prognostic factor.[32–34] As prognostic factors calculate imbalance measures, we collect 10 site-level characteristics that could affect both the treatment assignment process and our outcome measures, from site institutions' administrative records – the number of graduate students, % of HU students, % of female faculty, faculty's racial composition (% of White faculty and % of Asian faculty), faculty rank (% of assistant faculty and % of full professor), % of graduate foreign nationals, previous mentoring training (Y/N), and geographic region (Northeast, South, Midwest, West). Because we randomize at the site level and then recruit participants for that condition, we optimize balance only with site-level information, not individual-level information and outcomes of interest, which is one of our limitations.

Algorithm of Calculating the Measure of Imbalance

While calculating the measure of imbalance between two treatment groups was straightforward, the three-group randomization involves more complicated calculations. Among many alternatives, we apply a pairwise group comparison for our three-group randomization, proposed by Russell and colleagues.[34] Simply speaking, we assign a rank to given sites (clusters) regarding a variable of interest, and then use this rank matrix to calculate the difference between treatment groups. For categorical variables (e.g., previous mentoring training, geographic region), we use the number of sites for each stratum of each variable, which is most commonly used to calculate the imbalance scores for categorical variables.

AIM I Data Analysis

Using pre- and post-survey data, we will evaluate the comparative effects of CAM interventions on participants' scores on the study measures and their behavioral intentions. To account for a nested data structure of individual research mentors within university sites, we will use multilevel models (known as linear mixed-effect models) for our continuous outcomes.[35] Individual characteristics (e.g., demographics-including disaggregation by gender, faculty rank, previous mentoring experience) and cluster-level characteristics (e.g., those used in randomization) will be included in subsequent statistical analyses to further adjust for covariates not fully balanced in the experimental design. Using our longitudinal survey data, we will examine which factors facilitate or inhibit mentors' actual CAM behaviors after exposure to the CAM intervention through several statistical techniques. Multiple regressions and multivariate analyses of covariance will examine the predictive relationships between the three TPB

constructs (attitudes, norms, and control), behavioral intentions for CAM practices, and actual CAM behaviors in relation to mentors' demographics and CAM intervention treatment group. Finally, we will apply a structural equation modeling analysis to examine relationships among the measured constructs along with mentors' demographics, past experience, and institutional contexts.

AIM I Qualitative Interviews

Qualitative interviews are conducted with a sample of participants from each of the three trial arms 10-12 months after completion of the intervention. The broad goal of the interviews is to determine what was remembered from the intervention, reflections on how it changes thinking, and any actions taken which were closely linked to CAM participation. The time between training and interviewing purposefully allows time for mentors to use learned knowledge and skills so we could record behavioral changes. The CAM pilot data revealed that participants recall and can detail the intervention's impact 18-24 months later and, as such, we expect sufficient participant recollection at 10-12 months.

Interview Protocol

Interviews are conducted using a protocol designed by the study team. The interview question protocol covers pre-CAM experiences with mentoring, participant experiences in CAM, and any changes in thinking, behavior or actions that result from CAM participation. Changes related to mentoring are central to the questions, but the protocol also gathers data regarding changes in interactions with faculty peers and any other change they attribute to participation in CAM. The interviews are conducted in a semi-structured manner, allowing flexibility to follow up on and gather deeper insights by

asking clarifying questions. The interview protocol was field tested by one of the study co-investigators (RM) who has years of experience in bench research, biomedical research training, social science research, and the creation of CAM. All interviews are conducted by an experienced qualitative social science researcher (RJ) with deep knowledge of biomedical research training obtained through work on other research studies. Interviews are conducted over Zoom; audio files are uploaded to a secure server and transcribed by a professional service. To assess the transcription service quality, an initial small batch of interviews transcripts are compared to the audio recordings. The transcripts are uploaded as Word documents to Dedoose qualitative analysis software.

Interviewee Sampling

To achieve good sampling without unreasonable time and cost, we sample approximately 40% participants from each of the three trial arms distributed equally across the 33 sites. Given that racially and ethnically minoritized mentors are few in number at R1 institutions, we purposefully oversample minoritized participants to make sure impressions, benefits, and critiques from individuals of various racial-ethnic backgrounds are included. We sample to include faculty at different career stages and positions, from assistant professors to senior faculty appointed as deans. We sampled to include individuals from every gender. We recognize faculty who grew up in and/or were trained in other countries could have different views or experiences with mentoring than minoritized cultures in the U.S. For this reason, we also consciously sample to include participants whose first language is not English as a proxy for this demographic group.

About 10 months after completion of the training at a given university, the site host sends an email to all participants to inform them that a member of the study team may contact them for a short (less than 30 min) interview. An initial selection of about 50% of participants are then contacted and invited to a Zoom interview. After a week, those who have not responded are contacted a second time. If a sufficient number of interviewees have not been identified in this manner, additional participants are contacted until the target 40% is reached.

Interview Data Analysis

Three members of the qualitative research team (RM, RJ, FB) construct a general coding architecture with the interview protocol as a starting point. The study team is also interested in exploring how the Theory of Planned Behavior[17] might apply to the experiences and results of CAM participants, so codes to cover institutional norms, attitudes, sense of control, intended behavioral changes, and realized behavioral changes are included. The coding architecture is then applied to 90 interviews by a team member (FB) with 8 years of qualitative coding experience after initial calibration by the study team members who conduct the interviews. Passages of text that do not fit into the coding architecture as designed are also identified, and additional codes are developed if needed in this process. Next, the two members of the study team (RM, RJ) who conduct the interviews read the resultant code reports, one code at a time, to make minor adjustments through consensus. The same study members then read reports to identify more refined themes in each code and created sub-codes based on those themes. One of the interviewers (RJ) will carry out the remaining sub-coding and coding for a total of 181 interviews.

Study Measures and Data Collection

Survey Data

Table 2 summarizes what data are collected across the four survey periods, along with sample items, starting with the presurvey through the 12-month time period. The four survey periods are: baseline prior to any interventions (PRE), immediately following the intervention (POST), 6mos and 12mos.

We note that in the pre-survey, NRMN required measures such as demographic questions including race/ethnicity, gender identity, disability, parent/guardian education and degree completed as well as questions about disruptions in mentoring relationships due to COVID-19, and those related to social unrest related to racism after the murder of George Floyd. In the post-survey immediately following the intervention, we collect evaluation data on the iCAM module and for those that participate in the facilitated workshop, they are asked questions regarding the value of CAM workshop activities, timing, facilitation, and materials. We also administer a facilitator fidelity survey to have facilitators confirm which activities are completed and if there are any changes made to the curriculum for each workshop.

Qualtrics software at the University of Wisconsin-Madison is being used for all data collection. Most of the communication between the CAM study team and study participants is sent from the CAM study email address with the exception of the research manager (EE) and the graduate student assistant who occasionally send direct communications from their University of Wisconsin-Madison email addresses.

Cultural Diversity Awareness-Race/Ethnicity version. This validated scale assesses cultural diversity awareness related to race/ethnicity (CDA-R/E) in research

mentoring relationships informed by multicultural counseling theory and social cognitive theory.[36] 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 scale is not used in the current study. Responses are made on a 5-point Likert-type scale ranging from 1 (strongly disagree, never, not at all confident) to 5 (strongly agree, all the time, completely confident).

Mentoring Competency Assessment. This is a validated 26-item skills inventory that allows research mentors and mentees to self-evaluate six mentoring competencies: maintaining effective communication, aligning expectations, assessing understanding, addressing diversity, fostering independence, and promoting professional development.[37] To reduce response burden, Zhong et al.[38] created an 8-item 'short' version using exploratory factor analysis, confirmatory factor analysis and item response theory that still aligns with the original 6 mentoring competency domains. Given the length of our post-workshop survey, we opted to use this abbreviated version. Responses are made on a 7-point Likert-type scale where 1 = "not at all skilled" to 7 = "extremely skilled."

Department Climate. Hurtado and Carrasco[39] developed this survey to identify factors to assess biomedical science faculty members' perceptions about and behaviors related to cultural climate in their graduate training departments and programs. Responses to survey items were on a 5-point Likert scale with 1 for "strongly disagree" and 5 for "strongly agree." During the survey's development, all items were reviewed by the researchers across teams collaborating on the larger project, one who is an expert on campus climate research and another who is expert in cultural awareness training

among faculty in biomedical fields. Based on expert feedback in alignment with past literature, item wording was refined.

Theory of Planned Behavior Constructs. The Theory of Planned Behavior (TPB) includes five categories to evaluate behavioral change: attitudes, norms, control, behavior intentions and actual behaviors.[17] We include items to measure each of these categories. Attitudes and control are assessed with the CDA Attitudes and Confidence scales, respectively. Norms are measured with items in the Department Climate scale. We measure both injunctive norms, or norms perceived as valuable to others and pressure to conform to those norms (what one ought to do), and descriptive norms which capture an individual's beliefs about others' behavior (what is done). Items are included to measure perceptions about multiple stakeholders, including leadership, faculty peers, and trainees. We measure both intended and actual behavior change using the CDA Behavior scale in addition to items we created to assess likelihood of change. Finally, we include open-ended questions for participants to describe their intended and actual behavior change. (Not shown in Table 2 separately given that items span across multiple measures).

Culturally Aware Mentoring Skills. Individuals rate their perceived level of skill before and after the CAM training on four items in several areas including addressing issues of race/ethnicity when they arise in their mentoring relationships.[15] Responses are made on a 7-point Likert-type scale ranging from 1 = "not at all skilled" to 7 = "extremely skilled." The CAM skills items have not been factor analyzed; they are being used as individual items or as a composite score.

Rationale for timing of data collection

The timepoints of data collection on the outcome measures were determined based on our reasoning of a timeframe when changes in the measured construct might be evident. For example, we hypothesized that it is reasonable for attitudes and confidence related to CDA to change from prior to the intervention to immediately afterwards and thus collect survey data on the CDA measure in all four survey periods. Conversely, we did not expect that there would be detectable changes in participants' perceptions of their department climate prior to and just after the intervention and thus collect survey data on this measure at the 6mo and 12mo timepoints after baseline.

Process for sending surveys

All eligible participants are sent the presurvey to their self-reported email address. Participants are formally registered for the study by signing the informed consent form at the end of the presurvey. If the survey and signed informed consent are not completed, the survey is sent two additional times. If the presurvey still is not completed, they receive no further communications from the CAM study. Once registered, iCAM only participants (control group) are sent an email confirming their registration and are asked to reserve at least 3 hours on their calendar over the specified two-week intervention period to ensure availability to complete the self-paced intervention. All workshop participants are sent an email confirming their registration and reminding them of the workshop dates.

Intervention procedures

iCAM (control) interventions are scheduled for a two-week period starting on a Monday. To ensure that all participants have two full weeks to complete the intervention and mitigate delays due to technical difficulties, instructions to create a University of

Wisconsin-Madison Interprofessional Continuing Education Partnership (ICEP) account are sent to registered participants the Friday afternoon prior to the start of the intervention along with the iCAM access code. This site is where participants access the iCAM online module. Over the course of the two-week intervention period, participants receive three reminders with a summary of their iCAM completion status. The iCAM immediate post survey is sent to all registered participants on the Monday morning following the close of the intervention and is open for two weeks. Because we use an intent-to-treat approach, all surveys are sent to all registered participants regardless of their participation or engagement with the intervention unless the team receives a request to withdraw from the study.

Once registered for the study, participants receiving the CAM facilitated workshop receive a confirmation email with a calendar invite for the first session of the workshop, a link to the site-specific website where they have access to workshop activities and resources, and final pre-workshop tasks. The latter includes the same instructions iCAM participants receive to create a University of Wisconsin-Madison ICEP account. They also receive the iCAM access code with a request to complete all sections before the start of the workshop. Participants receive a PDF attachment of the Culture Box activity instructions and are asked to prepare a personal Culture Box to share with workshop participants during the first session. The study team sends an email one week before the first session of the workshop and another the day before that includes all the details mentioned above and the Zoom access information.

Participants receive an email between each session that includes the workshop zoom access information for the following session and the intercession reflection prompt

with a reminder to access and review content on the workshop website. The immediate post survey is sent in an email 15 minutes prior to the close of the workshop. The final slide of the workshop includes a reminder that the survey can be found in their email inbox. The facilitation team also makes a verbal request to complete the survey and, for anyone that is not able to find the survey, they are directed to contact the research manager (EE) for support. Participants that do not complete the immediate post, 6-month, or 12-month longitudinal survey are sent three reminder emails, all of which include a direct link to their personal survey. The final reminder is sent with a personal request from the research manager (EE). In cases where following the standard survey timeline means distributing surveys over a holiday or common vacation season, the survey collection period is extended an additional two weeks. The study team also makes phone calls to study participants at sites with a response rate lower than 70% of the registered participants.

AIM II Case Study

Central Purpose. Aim II seeks to understand the conditions for the broader organizational impact of the CAM training intervention in biomedical science PhD training programs. To this end, Aim II focuses on organizational dynamics regarding the CAM intervention, diversity, equity and inclusion (DEI) and mentoring activity within institutions. We employ an explanatory case study design to understand organizational change, specifically focusing on *how* new knowledge introduced into graduate programs results in broader change in diversity outcomes and department practices, and *how* and *why* CAM interventions catalyze program change for diversity and inclusion. It is important to note that case studies are best suited to understanding broader contexts as

well as specific organizational contexts that are part of the ecosystem of educational practice.

Using a multiple-embedded case study research design,[40] we aim to determine if CAM implementation in departments across institutions and across departments within institutions works similarly regarding inclusive science conditions for change and integration of CAM knowledge. By comparing departments and graduate programs across institutions in cross-case analysis, we understand the patterns between racial climate conditions, practices, institutional transformation with regard to diversity and mentoring. Departmental contexts and graduate programs are the main place for interaction among faculty peers and differences across these contexts helps to understand resistance to diversity and mentoring, and how faculty shape the environment for PhD trainees. Furthermore, this approach allows us to understand how campuses sustain successful student and faculty initiatives and institutionalize elements of CAM. Finally, we assess organizational impact by understanding what, if anything, changes in departments and programs due to CAM training and how CAM fits into the overall context of DEI work in schools of biomedical sciences and the universities.

Multiple Embedded Case Study Design. The AIM II team members (SH, AC) gather the following forms of evidence: participant surveys of the climate, archival records, document collection, semi-structured interviews, and campus observations. The embedded cases are illustrated in the institution's organizational levels: university site, school, department, graduate program, and CAM participants. A detailed description of the procedures for the case study method follows.

For Aim I, we administer the CAM intervention to biomedical science faculty at 33

R1 universities. For Aim II, we select one institution from this sample and a second institution not from this sample to gain a deeper understanding of organizational contexts. Participants at both institutions receive the three-session CAM intervention.

The team collects three forms of evidence before conducting interviews and campus visits from the selected institutions: IPEDS/graduate program data, survey data, and document collection. Utilizing IPEDS and graduate program data acquired from graduate program or DEI administrators, we ensure each institution had a Medical Center and Life Science biomedical PhD training programs and had at least 15 percent HU enrollment in each department. Additional rationale for site selection includes compelling DEI activity at the institutional and department level, which is determined through web scraping of online document collection. Web scraping gathers electronic documentation at multiple institutional levels including: mission, history, and strategy documents, campus demographics, general academic policies and requirements, campus and community news, and key contact names and profiles. This critical information allows us to build a timeline of events, contextualize CAM training in the institutional ecology, and corroborate data with additional sources of evidence to determine the validity of the information. We also incorporate the survey and outcome data described in the AIM I section from immediate post-data collection, 6-month post-data collection, and 12- month post data collection documenting responses to the department climate questions.

The Aim II team conducts virtual and in-person semi-structured interviews with members of the following organizational levels at both sites: institutional, school, department, graduate program, and CAM participants. Aim II team members have

developed semi-structured interview protocols that address questions in three key areas: diffusion of knowledge, CAM and organizational dynamics, and organizational change. We will conduct about 24 interviews at each site with administrators within their university's school or division (e.g., School of Medicine, Division of Biological Sciences), department chairs, graduate program staff, and CAM study participants at that site. All interviews except those conducted during each campus visit are conducted virtually on Zoom. They are audio-recorded and professionally transcribed. Aim II team members review and clean transcripts to ensure accuracy prior to the coding process.

Site visits occur after most virtual interviews are complete so that there is a holistic understanding of the organizational context from the previous forms of evidence before visiting the campus. Campus visits include follow-up interviews with key informants and remaining CAM participants, tours of biomedical spaces and facilities for observations, and collection of internal documents, including posters, pamphlets, and other archival records to contextualize DEI and mentoring activities within the various organizational levels.

AIM II Planned Analyses. Aim II team members will develop the codebook through deductive and inductive coding techniques. Inductive coding considers emerging dimensions organically stemming from the data and context, and deductive coding takes a theoretical approach to understand how theory appears within the data. The theoretical frameworks guiding Aim II data include the Campus Racial Climate framework[41] and Organizational Change frameworks[18] to understand how external and internal events motivate departments to participate in CAM training and how their involvement results in processes that lead to changes within the departments and

programs. Coding will be conducted using MAXQDA (2020) qualitative software to facilitate the development of themes and conduct cross-case analysis.

Discussion

Because research training is so completely dependent on mentorship, no real change is possible without a dramatic change in how faculty approach mentorship in general and mentorship across racial and ethnic differences in particular. The funding of NRMN fortuitously has enabled a critical window of opportunity for improving mentorship for HU trainees. Due in large part to the NRMN initiative, receptivity and opportunity to expand the reach of *Entering Mentoring*[42] is leading to a dramatic increase in training, coinciding with NIH and other research training organizations turning mentorship training into an expected norm, not an afterthought. Yet, even *Entering Mentoring* was not designed to get into the deep conversations, self-awareness and expanded skills needed by faculty to effectively mentor students from HU groups.[43] This set the backdrop of need, interest and opportunity to create CAM.

The CAM intervention truly is unique in that it takes a scientific approach to diversity interventions in the biomedical sciences by incorporating four curricular design elements and targeting measurable participant outcomes identified as best practices for such interventions.[14] The outcomes from this current randomized comparative study will provide extensive new insights into CAM as an intervention and how it impacts faculty self- and other perceptions, their skills and their behaviors, and how it may impact the culture and practices of graduate departments and programs.

But the goals of the study, even at the individual faculty and department or

training program levels, are actually bigger. The study is designed to provide new insights into how other interventions aimed at advancing evidence-based, culturally responsive practices might be crafted. The CAM intervention is a starting point, and future interventions will need to help faculty go even farther into refining their culturally responsive mentorship effectiveness. We assert that many of the theoretical frameworks and underpinnings in CAM are adaptable to other interventions promoting cross-cultural understanding and communication. We hope that providing this detailed study protocol will enable others to apply them to related efforts.

Inherent limitations in this study include self-reported outcomes, self-selection bias and period effects. For example, faculty who volunteer to participate in the study may already be motivated to improve their mentoring effectiveness in general. Further, our study occurs since the summer of 2020's race-related social upheaval, that included numerous historical events like the murder of George Floyd and the trial of Derek Chauvin. One outcome of these historical events is that many faculty have been subsequently exposed to increased DEI-related professional development opportunities at their universities. Engagement in such opportunities can augment or compromise the intended effects of the CAM intervention, depending on the quality and/or quantity of participants' experiences with other professional development. Finally, because we are examining longitudinal effects of an intervention over 12 months, we cannot discount the effect of natural maturation such that people grow and change over time regardless of participating in the study's treatment.

Authors Contributions

EE contributed to data curation, project administration, resources, supervision, writing of the manuscript. ABW contributed to conceptualization of the study, data curation, funding acquisition, investigation,

methodology, resources, writing of the manuscript. SCH contributed to data curation, investigation, methodology, project administration, writing of the manuscript. RM contributed to conceptualization, funding acquisition, investigation, methodology, and writing of the manuscript. RJ contributed to data curation, methodology, and writing of the manuscript. SH contributed to conceptualization, data curation, funding acquisition, investigation, methodology, and writing the manuscript. AC contributed to data curation, methodology, and writing of the manuscript. YL contributed to conceptualization, methodology software, and writing of the manuscript.

Acknowledgements
We thank Ana Gomez and Bernard Reyes for their intellectual support in the early phases of the project.

Figure 1: ***Multi-Phase Mixed Method Study Design***

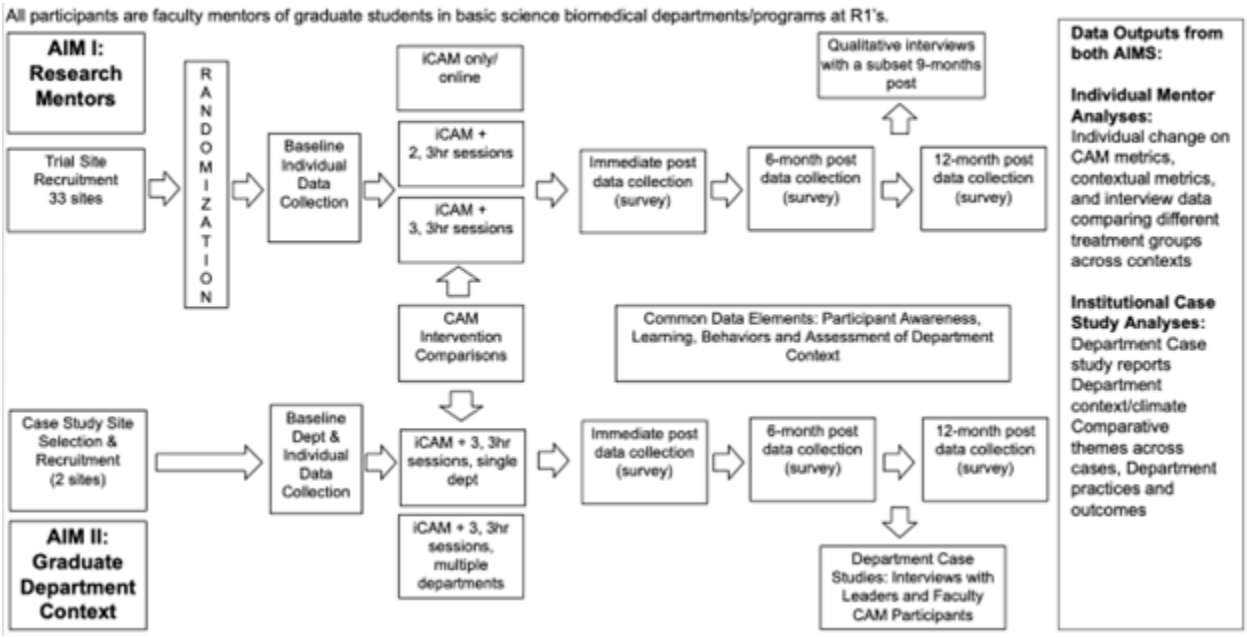

Table 1: ***Culturally Aware Mentoring Intervention Comparison***

Activity	iCAM + 3-session CAM	iCAM + 2-session CAM	iCAM only
Part 1: Race and Privilege	x	x	x
Part 2: The experiences of scientists from historically underrepresented groups	x	x	x
Part 3: The realities of cultural diversity in the sciences	x	x	x
Part 4: The role of culturally aware mentoring in trainee outcomes	x	x	x
iCAM debrief	x	x	
Intrapersonal			

Culture Box Activity	x	x	
Personal Reflection #1: Racial Identity	x		
Intersession Reflection #1	x	x	
Interpersonal			
Video: Tale of O	x	x	
Personal Reflection #2: Culture on Your Research Group	x	x	
Seeing Color	x		
Research on color blindness/color evasiveness	x	x	
I Fit the Description	x	x	
Intersession Reflection #2	x		
Skill Building			
Principles for Culturally Aware Mentoring Practices	x	x	
Research on Broaching Challenging Conversations	x	x	
Personal Reflection #3: Race/Ethnicity in the Mentoring Relationship	x		
Case Scenario: Trainee Differences	x	x	
Role Play: Trainee Differences (2 rounds)	x	x	
Action Plan: Next Steps in Your CAM Journey	x	x	

Table 2: ***Culturally Aware Mentoring Measures***

Measure/scales	**Pre**	**Immediate Post**	**6 Month**	**12 Month**
Facilitator Fidelity Survey				
Please note which activities were implemented during session x of the workshop:		x		
Please note any changes or adjustments you made to activities implemented during session x		x		
Please note any extraordinary circumstances that occurred during the workshop.		x		
Participant Surveys				
Culturally Aware Mentoring Evaluation				
Example: Please indicate how much you disagree or agree with each of the following statements: The facilitators were knowledgeable about the workshop content.		x		
Temporal/cultural context questions *How has the social unrest regarding systemic racism affected your mentoring relationships with graduate students?* *How much has the social unrest regarding systemic racism influenced your interest in the CAM workshop?*	x		x	x
Demographics *Did any of the adults who raised you (i.e. parent, stepparent, guardian), complete a bachelor's degree?*	x		x	x
Cultural Diversity Awareness-Race/Ethnicity version (CDA-R/E)				

Attitudes: *My racial/ethnic identity is relevant to my research mentoring relationships.*	x	x	x	x
Confidence*: Please rate your response to the following question from not at all confident to completely confident. Discuss with mentees how it feels to be a minority in science.*	x	x	x	x
Behaviors*: Please indicate how frequently each of the following has occurred in your research mentoring relationships during the past x months, including any current research mentoring relationship. I reflected upon how the research experience might differ for mentees from different racial/ethnic groups.*	x		x	x
Mentoring Competency Assessment (short)				
Rate how skilled you feel you were BEFORE/NOW for each of the following: 1=not at all skilled; 4=moderately skilled; 7=extremely skilled		x		
Taking into account the biases and prejudices you bring to your mentor/mentee relationship		x		
Understanding your impact as a role model		x		
Department Climate				
Faculty in my PhD program actively support graduate students from underrepresented racial/ethnic groups	x		x	x
I have initiated discussions about the importance of diversity in mentoring with faculty peers.	x		x	x
Culturally Aware Mentoring skills				
*Please rate how skilled you feel you were **before** the workshop and how skilled you feel you are **now** in each of the following areas. Going outside of my comfort zone to help mentees feel included in the training environment*		x		

References

1. National Academies of Sciences, Engineering, and Medicine. *The Science of Effective Mentorship in STEMM*. (Dahlberg ML, Byars-Winston A, eds.). National Academies Press; 2019. Accessed May 19, 2021. http://www.ncbi.nlm.nih.gov/books/NBK552772/

2. Acker J. Inequality regimes: Gender, class, and race in organizations. *Gend Soc*. 2006;20(4):441-464. doi:10.1177/0891243206289499

3. Valantine HA, Collins FS. National Institutes of Health addresses the science of diversity. *Proc Natl Acad Sci*. 2015;112(40):12240-12242. doi:10.1073/pnas.1515612112

4. Stanley CA. Coloring the academic landscape: Faculty of color breaking the silence in predominantly white colleges and universities. *Am Educ Res J*. 2006;43(4):701-736.

5. Malone KR, Barabino G. Narrations of race in STEM research settings: Identity formation and its discontents. *Sci Educ*. 2009;93(3):485-510. doi:10.1002/sce.20307

6. Milkman KL, Akinola M, Chugh D. Temporal distance and discrimination: An audit study in academia. *Psychol Sci*. 2012;23(7):710-717. doi:10.1177/0956797611434539

7. Sowell R, Allum J, Okahana H. *Doctoral initiative on minority attrition and completion*. Wash. DC: Council of Graduate Schools. 2015.

8. Ginther DK, Kahn S, Schaffer WT. Gender, race/ethnicity, and National Institutes of Health R01 research awards: Is there evidence of a double bind for women of color? *Acad Med J Assoc Am Med Coll*. 2016;91(8):1098-1107. doi:10.1097/ACM.0000000000001278

9. National Academies of Sciences, Engineering, and Medicine. *Graduate STEM Education for the 21st Century*. (Leshner A, Scherer L, eds.). National Academies Press; 2018. doi:10.17226/25038

10. Gibbs Jr. KD, McGready J, Bennett JC, Griffin K. Biomedical science Ph.D. career interest patterns by race/ethnicity and gender. *PLOS ONE*. 2014;9(12):e114736. doi:10.1371/journal.pone.0114736

11. Carnes M, Devine PG, Manwell LB, et al. Effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Acad Med J Assoc Am Med Coll*. 2015;90(2):221-230. doi:10.1097/ACM.0000000000000552

12. Capers Q, Clinchot D, McDougle L, Greenwald AG. Implicit racial bias in medical school admissions. *Acad Med J Assoc Am Med Coll*. 2017;92(3):365-369. doi:10.1097/ACM.0000000000001388

13. Wear D, Zarconi J, Aultman JM, Chyatte MR, Kumagai AK. Remembering Freddie Gray: Medical education for social justice. *Acad Med J Assoc Am Med Coll*. 2017;92(3):312-317. doi:10.1097/ACM.0000000000001355

14. Moss-Racusin CA, van der Toorn J, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. Scientific diversity interventions. *Science*. 2014;343(6171):615-616. doi:10.1126/science.1245936

15. Byars-Winston A, Womack VY, Butz AR, et al. Pilot study of an intervention to increase cultural awareness in research mentoring: Implications for diversifying the scientific workforce. *J Clin Transl Sci*. 2018;2(2):86-94. doi:10.1017/cts.2018.25

16. Womack VY, Wood CV, House SC, et al. Culturally aware mentorship: Lasting impacts of a novel intervention on academic administrators and faculty. *PLOS ONE*. 2020;15(8):e0236983. doi:10.1371/journal.pone.0236983

17. Fishbein M, Ajzen I. *Predicting and changing behavior: The reasoned action approach*. Taylor & Francis; 2011.

18. Hurtado S, White-Lewis D, Norris K. Advancing inclusive science and systemic change: The convergence of national aims and institutional goals in implementing and assessing biomedical science training. *BMC Proc*. 2017;11(12):17. doi:10.1186/s12919-017-0086-5

19. McDaniels M, Pfund C, Barnicle K. Creating dynamic learning communities in synchronous online courses: One approach from the Center for the Integration of Research, Teaching and Learning (CIRTL). *Online Learn*. 2016;20(1):110-129.

20. Rogers J, Gong X, Byars-Winston A, et al. Comparing the outcomes of face-to-face and synchronous online research mentor training using propensity score matching. *CBE Life Sci Educ*. 2022;21:ar62. doi:10.1187/cbe.21-12-0332

21. House SC, McDaniels M, Spencer KC, Utzerath E, Pfund C. The NRMN Master Facilitators Initiative: Building a community of practice to broaden program implementation. *Underst Interv.* Published online 2020.

22. House SC, Byars-Winston A, Zárate S, et al. Guiding principles for culturally responsive facilitation: Lessons learned from delivering culturally aware mentor training to STEMM faculty. *J Divers High Educ*. Published online 2023. doi:10.1037/dhe0000476

23. Weber-Main AM, Shanedling J, Kaizer AM, Connett J, Lamere M, El-Fakahany EE. A randomized controlled pilot study of the University of Minnesota mentoring excellence training academy: A hybrid learning approach to research mentor training. *J Clin Transl Sci*. 2019;3(4):152-164. doi:10.1017/cts.2019.368

24. Pfund C, House SC, Asquith P, et al. Training mentors of clinical and translational research scholars: A randomized controlled trial. *Acad Med J Assoc Am Med Coll*. 2014;89(5):774-782. doi:10.1097/ACM.0000000000000218

25. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. L. Erlbaum Associates; 1988.

26. Hoop E de, Teerenstra S, Gaal BGI van, Moerbeek M, Borm GF. The “best balance” allocation led to optimal balance in cluster-controlled trials. *J Clin Epidemiol*. 2012;65(2):132-137. doi:10.1016/j.jclinepi.2011.05.006

27. Perry M, Faes M, Reelick MF, Rikkert MGMO, Borm GF. Studywise minimization: A treatment allocation method that improves balance among treatment groups and makes allocation unpredictable. *J Clin Epidemiol*. 2010;63(10):1118-1122. doi:10.1016/j.jclinepi.2009.11.014

28. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. *Trials*. 2012;13(1):120. doi:10.1186/1745-6215-13-120

29. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975;31(1):103-115. doi:10.2307/2529712

30. Taves DR. Minimization: A new method of assigning patients to treatment and control groups. 1974. In Clinical Pharmacology & Therapeutics - Wiley Online Library. Accessed December 15, 2023. https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt1974155443

31. Saghaei M. An overview of randomization and minimization programs for randomized clinical trials. *J Med Signals Sens*. 2011;1(1):55-61.

32. Chen LH, Lee WC. Two-way minimization: A novel treatment allocation method for small trials. *PLOS ONE*. 2011;6(12):e28604. doi:10.1371/journal.pone.0028604

33. Raab GM, Butcher I. Balance in cluster randomized trials. *Stat Med*. 2001;20(3):351-365. doi:10.1002/1097-0258(20010215)20:3<351::AID-SIM797>3.0.CO;2-C

34. Russell D, Hoare ZSJ, Whitaker Rh, Whitaker CJ, Russell IT. Generalized method for adaptive randomization in clinical trials. *Stat Med*. 2011;30(9):922-934. doi:10.1002/sim.4175

35. Bryk AS, Raudenbush SW. *Hierarchical linear models: Applications and data analysis methods*. Sage Publications, Inc; 1992:xvi, 265.

36. Byars-Winston A, Butz AR. Measuring research mentors’ cultural diversity awareness for race/ethnicity in STEM: Validity evidence for a new scale. *CBE—Life Sci Educ*. 2021;20(2):ar15. doi:10.1187/cbe.19-06-0127

37. Fleming M, House S, Hanson VS, et al. The mentoring competency assessment: Validation of a new instrument to evaluate skills of research mentors. *Acad Med J Assoc Am Med Coll*. 2013;88(7):1002-1008. doi:10.1097/ACM.0b013e318295e298

38. Zhong S, Maccalla NMG, Jeon M. *Short-Form of the Mentoring Competency Assessment (MCA-Short)*. National Institutes of Health Diversity Program Consortium; 2020.

39. Hurtado S, Reyes B, Carrasco A. The scholarship on department and graduate program racial climates in STEMM fields. Presented at: AERA 2022; April 14, 2022. https://seis.ucla.edu/news/ucla-at-aera-2022--a-preview

40. Yin RK. *Case study research and applications: Design and methods*. SAGE Publications; 2017.

41. Hurtado S. The campus racial climate. *J High Educ*. 1992;63(5):539-569. doi:10.1080/00221546.1992.11778388

42. Pfund C, Branshaw J, Handelsman J. *Entering Mentoring*, 1st Edition. Macmillan Learning US. Published 2015. Accessed January 4, 2024. https://www.macmillanlearning.com/college/us/product/Entering-Mentoring/p/1464184909

43. House SC, Spencer KC, Pfund C. Understanding how diversity training impacts faculty mentors' awareness and behavior. *Int J Mentor Coach Educ*. 2018;7(1):72-86. doi:10.1108/IJMCE-03-2017-0020