5-7-2020

Mixed-Reality Simulations to Build Capacity for Advocating for Diversity, Equity, and Inclusion in Geosciences

Jason A. Chen

M. Shane Tutwiler

Jerlando F. L. Jackson

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

American Psychological Association

NADOHE
National Association of Diversity Officers in Higher Education

Journal of Diversity in Higher Education

ISSN: 1938-8926

http://dx.doi.org/10.1037/dhe0000190

AQ: 1

Mixed-Reality Simulations to Build Capacity for Advocating for Diversity, Equity, and Inclusion in the Geosciences

AQ: au
AQ: 2

Jason A. Chen
William & Mary School of Education

M. Shane Tutwiler
University of Rhode Island

Jerlando F. L. Jackson
University of Wisconsin, Madison

We report on data collected at 3 time points during a 1-year intervention designed to teach a purposive sample of geoscience faculty members (n = 29) from 27 universities throughout the United States how to identify and address issues related to diversity, equity, and inclusion in their departments. For the intervention we used mixed-reality simulations to help participants practice specific skills to address common situations in geoscience departments. The intervention also included an intensive 3-day workshop and 3 journal clubs. Using a Bayesian analytical approach we explored: (a) general trends in participants' self- and collective efficacy for identifying and addressing diversity, equity, and inclusion over a 1-year period; (b) relationships between self-efficacy and collective efficacy; and (c) demographic factors that explain variation in self- and collective efficacy. Results showed that self- and collective efficacy rose sharply from preintervention to 5 months after beginning. Although both self- and collective efficacy retreated toward baseline at the 1-year mark, only 1-year self-efficacy was still credibly higher than preintervention. Also, preintervention self-efficacy predicted 5-month collective efficacy. Efficacy beliefs varied as a function of race/ethnicity. Only collective efficacy varied as a function of academic rank. We discuss these findings in relation to social–cognitive theory and the literature regarding the use of digital learning environments to address diversity, equity, and inclusion.

Keywords: technology, simulations, social–cognitive theory, Bayesian analysis, efficacy beliefs

AQ: 3 *Supplemental materials:* http://dx.doi.org/10.1037/dhe0000190.supp

There is growing evidence that diversity can improve a group's productivity and creative problem-solving (Page, 2008). In general, diverse scientific teams publish in higher quality journals, and their articles get cited at higher rates than do their more homogenous peer groups (Freeman & Huang, 2015). In the corporate world, companies that have a more diverse workforce and are more inclusive yield higher profits and are more highly motivated than their more homogenous peer companies (Herring, 2009; Forbes, 2011; Hunt, Layton, & Prince, 2015). From a quality-of-life perspective, diverse and inclusive workplaces also tend to have employees who are happier, less likely to leave the organization for another one, and are more mentally and physically healthy (Goffee & Jones, 2013; Hitlan, Cliffton, & DeSoto, 2006; Nadal, 2011). Yet, in the field of geosciences only 6% of all doctorate degrees in 2016 were conferred to underrepresented minority students (i.e., defined by the National Science Foundation as those who are not White non-Hispanic or Asian non-Hispanic), which is the lowest proportion among all STEM fields (Bernard & Cooperdock, 2018). To put that number in perspective, 31% of the American population comes from underrepresented minorities. Also alarming is the fact that, for over 40 years, despite efforts to increase diversity in the STEM fields (geosciences included), the proportion of underrepresented minorities has not changed (Bernard & Cooperdock, 2018).

Diversity Training

Amid this backdrop, organizations have turned to diversity training as a way to make workplaces more diverse, equitable, and inclusive (Dobbin & Kalev, 2018). Unfortunately, there is mounting evidence that the typical diversity training for employees, including university faculty members, is largely ineffective at changing attitudes or behaviors (Bezrukova, Jehn, & Spell, 2012; Moss-Racusin et al., 2014). One possible reason for this is because there is low engagement in this type of training (Williams, 2013), and the training provided does not help participants take on specific behaviors to counter implicit and explicit biases or to push against institutional inertia (Bezrukova et al., 2012; Moss-Racusin et al., 2014). These traditional diversity training modules are

Jason A. Chen, William & Mary School of Education; M. Shane Tutwiler, School of Education, University of Rhode Island; Jerlando F. L. Jackson, School of Education, University of Wisconsin, Madison.

This work was supported by a grant from the National Science Foundation under Grant ICER-1645444.

Correspondence concerning this article should be addressed to Jason A. Chen, William & Mary School of Education, P.O. Box 8795, Williamsburg, VA 23187-8795. E-mail: jachen@email.wm.edu

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

focused mostly on compliance—what individuals should avoid so as to protect themselves and their organizations from financial and legal harm and embarrassment (Bezrukova et al., 2012; Moss-Racusin et al., 2014).

Given the above-mentioned problems of a lack of diversity in geosciences coupled with the poor history of diversity training, the overall goal of the project is to apply a novel technological tool (i.e., mixed-reality simulations) to teach a cohort of geoscientists how to champion efforts to create more diverse, equitable, and inclusive geoscience departments in their universities.

Technology-Enabled Diversity Programs

In addition to analog diversity training programs, computer-based diversity programs have been gaining some attention. For example, DiversityEdu is a popular and widely available computer-based program that many institutions have implemented with university students, staff, and faculty. However, most researchers have focused on the effectiveness of computer-based diversity programs for students (Goldstein Hode, Behm-Morawitz, & Hays, 2018). Little research has explored the effectiveness of computer-based programs for faculty to address diversity, equity, and inclusion issues. However, many of the computer-based programs mostly digitize traditional content and training—offering little to no opportunity for participants to actively learn behaviors specific to promoting diversity, equity, and inclusion. Thus, these programs digitize the very problems already mentioned regarding traditional diversity training, and likely do little to resolve the problems.

Prejudice Habit-Breaking Intervention

Despite the underwhelming results of traditional diversity programs, whether in-person or technology-mediated, high quality professional development focused on prejudice- and bias-reduction can be effective, as demonstrated by the research team behind the prejudice habit-breaking interventions (e.g., Carnes et al., 2015, 2012; Devine, Forscher, Austin, & Cox, 2012; Forscher, Mitamura, Dix, Cox, & Devine, 2017). These researchers found that their prejudice habit-breaking intervention raised awareness and personal concern about discrimination as well as an increased tendency to label biases as wrong. Published papers regarding the prejudice habit-breaking intervention, including experimental studies, have been impressive and promising.

In light of the problems mentioned earlier, and given the only published intervention specifically targeting prejudicial behaviors (e.g., the prejudice habit-breaking intervention) the first author led the creation of the GeoDES project (Geoscience Diversity Experiential Simulations, which is a National Science Foundation-funded project) to take advantage of the affordances of mixed-reality simulations (i.e., simulations in which human conversational intuition combines with artificial intelligence to produce hyper-realistic scenarios) for learning and motivation. However, because interventions must be based on sound theory, we next describe the theoretical framework on which our project is based, and how the technology was designed in-line with theory to address the problem of teaching university faculty how to create more diverse, equitable, and inclusive work environments, and to direct their motivational resources toward such behaviors.

AQ: 4

Social–Cognitive Theory as a Guide for Designing Computer-Based Simulations

In designing our simulations, we drew from the teacher professional development literature. Hamre et al. (2012) concluded that, "interventions that primarily target beliefs and knowledge may have limited impacts on teachers' practice unless they directly focus on *practice* [emphasis added]" (p. 114). That is, helping people develop the behaviors and habits that are consistent with adaptive beliefs makes it more likely that they will not only change behaviors for the long term, but also will correspondingly change their beliefs. When it comes to advocating for diversity within geoscience settings, there are numerous complex behaviors that individuals need to deploy with considerable political and social tact to effect change. Missteps in such situations could prove embarrassing and put the person resisting institutional inertia in jeopardy. Therefore, individuals who put themselves in such a situation require a robust sense of efficacy (Bandura, 1997) to enact specific behaviors and choose specific words to counteract implicitly and explicitly prejudiced practices and structures.

Self-Efficacy and Long-Term Patterns of Change

Bandura (1997) defined self-efficacy as the "belief in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). Because self-efficacy is an important outcome in many interventions, our GeoDES project team focused on long-term changes in self-efficacy to identify and address issues related to diversity, equity, and inclusion. A small but growing body of literature regarding long-term self-efficacy changes in diversity professional development informed our work. For example, in a recent study, Goldstein Hode and colleagues (2018) reported on the effectiveness of an online diversity course for faculty and staff. They reported results on three self-efficacy items regarding the following tasks: (a) speaking up when they hear a colleague make disparaging jokes; (b) knowing what to do when they witness discrimination or harassment; and (c) knowing what concrete steps to take to make their university more inclusive. Compared to preintervention, participants reported being more self-efficacious to perform all three tasks. However, like many studies reporting self-efficacy changes over time, Goldstein Hode et al. reported pre–post changes over a relatively short period of time (4 weeks). In addition, Goldstein Hode et al. acknowledged that their self-efficacy scale was not sufficiently reliable, making the results on self-efficacy suitable mostly for preliminary exploration.

Some researchers have explored long-term changes in self-efficacy by going beyond just two time points. In one study, Combs and Luthans (2007) found that diversity training that focuses on bolstering participants' self-efficacy led to greater gains in self-efficacy as well as stronger intentions to act in ways that support diversity, equity, and inclusivity. The authors also found that diversity self-efficacy mediated the effect of the intervention on intentions to act in diversity-supportive ways. This effect held up even 1 year after the intervention.

Journeying outside of studies that deal with diversity, equity, and inclusion, Hoy and Spero (2005) reported that preservice teachers enrolled in a teacher preparation program became more self-efficacious from the start of the teacher preparation program

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

to the end of their program 1 year later. Hoy and Spero then followed up with these teachers again another year later when they had finished their first year of full-time teaching. They found that self-efficacy increased by the end of the teacher preparation program, but then retreated back toward baseline by the end of their first full year of teaching. However, participants were still more self-efficacious compared to when they first started the teacher preparation program. Perhaps, when met with the reality of having to teach their own students, and to take full responsibility of a class, these teachers became a bit more realistic in what they could and could not do successfully.

Collective Efficacy

Social–cognitive theorists contend that individuals do not function in a vacuum—they must interact with others, and are subjected to the behaviors and cultural norms created by a group (Bandura, 1986). Redressing issues related to diversity, equity, and inclusion is therefore a group effort requiring coordinated effort from teams of decision makers (Dobbin & Kalev, 2018). Following this line of reasoning, faculty members within a department must work collaboratively to achieve what they cannot achieve as an individual. Social–cognitive theorists extend the concept of self-efficacy beyond the individual to a group through the concept of collective efficacy, which can be defined as a group's shared belief in its collective power to achieve desired results (Bandura, 1997). Given the importance of collective efficacy in creating institutional change, we were curious about whether the GeoDES program could develop participants' confidence in working together with their department to create more diverse, equitable, and inclusive workplaces. However, the GeoDES project team focused on equipping individuals with the confidence to take on challenges related to diversity, equity, and inclusion rather than on equipping whole departments to do this work. For this reason we also wondered if developing participants' self-efficacy to identify and redress prejudices would also contribute to developing their confidence in being able to work together with their departments to create a more diverse, equitable, and inclusive workplace, which, as we discuss next, is what the empirical work of Fernández-Ballesteros, Diez-Nicolàs, Caprara, Barbaranelli, and Bandura (2002) suggests.

Relationship Between Self- and Collective Efficacy

There is reason to believe that people who have strong self-efficacy in a particular class of tasks may also have strong collective efficacy in working together with a group to accomplish a group goal. Fernández-Ballesteros et al. (2002) collected data from a nationally representative sample of 1,214 participants in Spain. They found that people's beliefs about society's ability to make positive social change (i.e., collective social efficacy) is partly the product of a robust belief in one's own ability to manage the demands of their own daily life (i.e., perceived personal efficacy) as well as their ability to contribute individually to the improvement of societal problems (i.e., individual social efficacy). They concluded that self-efficacy predicts collective efficacy. In fact, they also tested the alternative model in which collective efficacy predicts self-efficacy, and found weak support for this.

But "confidence in society's ability to make positive changes" does not intuitively seem to fit the definition of collective efficacy, which is the belief in one's capability to work together with others to accomplish a collective goal (Bandura, 1997, 2000; Lent, Schmidt, & Schmidt, 2006). Zooming in to a more context-specific setting dealing with group functioning, Lent et al. (2006) reported on a sample of undergraduate engineering students working together in teams to solve challenges they would face in real-life work situations. They found that individuals' confidence in their ability to cope with the types of obstacles that engineering students typically faced in their program (i.e., self-efficacy) predicted their confidence in the team's ability to work together in accomplishing tasks (i.e., collective efficacy). Therefore, when researchers investigate people's beliefs about being able to work together within a group to achieve specific outcomes, self-efficacy still predicts collective efficacy.

Using Immersive Technologies to Build Efficacy Beliefs

Advancing diversity, equity, and inclusion goals in higher education is a heavy lift, but it is possible that innovative technologies can help. As we discuss later, our mixed-reality simulations combined artificial intelligence with human conversational intuition to create a highly immersive and authentic experience that geoscience faculty members could realistically find themselves. These simulations were designed so that participants could develop their skills in navigating socially and politically tricky situations such as advocating for a job candidate who is not given the benefit of the doubt during a search process for a new tenure-track position in the geosciences department, while other candidates are given the benefit of the doubt. Although the simulations were a key component of the intervention, there were also other aspects of the intervention, which we describe next in the context of the whole project.

Overview of GeoDES Project

GeoDES is one of five pilot research projects funded through the National Science Foundation's GOLD program (Geoscience Opportunities for Leadership Development). One key goal of the GOLD program was to bring together researchers from diverse scientific backgrounds and perspectives to generate new approaches to develop leadership for broadening participation in the geosciences. The GeoDES team brings together research expertise from educational psychology, geosciences, and workforce diversity in higher education contexts.

The GeoDES team adapted research-based methods from teacher professional development (PD) and applied them to PD in equity and inclusion for geoscientists. The program consisted of three components (an illustration of the timeline of activities involved in the GeoDES project is available in the online study online supplementary material). First, through a 3-day in-person workshop, the GeoDES team provided a cohort of 29 geoscientists PD to develop their (a) knowledge of social justice issues in geosciences; (b) bystander intervention skills; and (c) leadership skills for targeting exclusionary gatekeeping decisions in university departments. This workshop featured two speakers from our project team who have expertise in diversity, equity, and inclusion in higher education.

The second component consisted of three simulations, which were built in collaboration with the technology company Mursion

Inc., to combine human conversational intuition with artificial intelligence. The purpose of this human-in-the-loop architecture is to make the scenarios so authentic and realistic that our participants would willingly "suspend their disbelief" (Dede, 2009) enough to engage in a difficult conversation (an illustration of this architecture is available in the online study online supplementary material). Participants were able to use the simulations in the comfort of their own home using their own laptop or desktop computer.

In addition, because self- and collective efficacy were important outcomes, throughout the simulation design the GeoDES team attended to Bandura's (1997) four hypothesized sources of efficacy beliefs: (a) interpreted successes of past performance (i.e., *mastery experiences*); (b) *vicarious experiences* of watching others perform a task; (c) the verbal and nonverbal feedback that trusted others provide (i.e., *social persuasions*); and (d) *physiological and affective states* such as cheerfulness, anxiety, or enjoyment.

By engaging participants in a simulated experience where they have to speak up and act within the simulated social situation, the GeoDES project team designed the simulation with mastery experiences in mind. The simulations were also designed with vicarious experiences in mind because all simulations are recorded, and because the GeoDES team encouraged all participants to view and then discuss during the journal clubs their performances in the simulation (although we cannot guarantee that everyone did watch their videos because it was not required). Finally, immediately after the simulation, participants were able to debrief with the "host avatar" to discuss what went well and what could have been done better. In this way, the simulations were designed with social persuasions in mind.

Participants engaged in the first simulation during the 3-day workshop. The second simulation took place within a month after the 3-day workshop. The third simulation was completed by October 2018. For Simulation 1 participants had to identify and "call in" microaggressions during a meeting with a White male department chair and an African American female colleague. "Calling in" allows the participant to bring awareness tactfully to the department chair that he said something insensitive. The focus here is on tactfully doing this so as to bring the department chair into the conversation without his acting defensively (i.e., "generating more light than heat").

For Simulation 2, participants advocated for a fictional, Latina job candidate who did not possess the "cultural assets" that the search committee values (e.g., PhD from an elite university, an advisor whom the committee members recognize, and publications in journals that the committee members prefer), even when the Latina candidate's record of achievement is comparable to others. In Simulation 3 participants made a compelling case for the benefits of diversity, and advocated for change in the department's annual merit review process, which ties faculty pay raises to activities related to diversity, equity, and inclusion. Participants had to tactfully manage push-back from resistant department members.

Because authenticity was an important goal in designing these simulations, the GeoDES team designed them by first interviewing over a dozen geoscientists about what experiences of social exclusion, prejudice, and microaggressions they have faced. After finding common themes that surfaced in these interviews, GeoDES team members began the process of writing scripts for the simulations in collaboration with Mursion, Inc.'s staff of professionally trained actors. Once these initial scripts were completed, the GeoDES project also consulted with a university-based theater troupe that specializes in interactive theater in university settings. This theater troupe often performs interactive theater to help university faculty and staff engage in difficult dialogues about issues of diversity, equity, and inclusion. Our final revisions were made in response to the interactive theater troupe's feedback.

Finally, the third component of the GeoDES project included three journal clubs where our program staff led virtual real-time discussions about readings that would further participants' learning about diversity, equity, and inclusion. These meetings were also spent discussing how they could apply what they have learned thus far to their specific institutional context. Project members who participated in these journal clubs included a faculty member in geosciences who has been engaged in diversity, equity, and inclusion work at his own institution. Another project member is the chief diversity officer of a major professional organization in geosciences. Finally, another project member has expertise in education, psychology, and technological innovations for learning. Readings included popular press articles that dealt with diversity, equity, and inclusion (e.g., West, 2017) as well as articles from science and research journals (e.g., Kalev, Dobbin, & Kelly, 2006; Mitchneck, Smith, & Latimer, 2016).

Research Questions and Hypotheses

Given the purposes and theoretical framework that undergird our study, the following research questions (RQ) and hypotheses (H) guided the research inquiry:

RQ1: What general trends in participants' self- and collective efficacy were evident from preintervention (Time 0 [T0]) to roughly 5 months postintervention (Time 2 [T2]) and then to roughly 1 year postintervention (Time 3 [T3])? AQ: 5

> *Hypothesis 1:* Given prior work showing that beliefs about efficacy for a task tend to spike immediately after an intervention and then tend to retreat toward baseline levels after a longer period of time (Hoy & Spero, 2005), we hypothesized a similar pattern of spiking and retreating toward baseline levels in our participants.

RQ2: To what degree does initial (T0) and 5-month postintervention (T2) self-efficacy vary by initial (T0) collective efficacy? To what degree does T0 and T2 collective efficacy vary by T0 self-efficacy?

> *Hypothesis 2:* Based on prior work investigating self- and collective efficacy we hypothesized that individuals' self-efficacy to identify prejudices and intervene appropriately would have an effect on their beliefs about whether they could work collaboratively within their own institution's departments to identify and redress prejudices (i.e., collective efficacy; Bandura, 2001; Fernández-Ballesteros et al., 2002; Knoblauch & Hoy, 2008). Given this literature, we also hypothesized that collective efficacy would not have an effect on self-efficacy.

RQ3: What, if any, demographic factors explain variation in the rise in efficacy beliefs (collective or self) from preintervention (T0) to five months postintervention (T2)?

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

Hypothesis 3: Because addressing issues related to diversity, equity, and inclusion has much to do with power and access to resources, and given that race, gender, and faculty rank all have power implications, we hypothesized that changes in efficacy beliefs would vary as a function of these demographic variables. However, we do not forward any directional hypotheses, as this was exploratory.

Method

The research described here was approved by the lead author's institutional review board. To answer our four research questions the GeoDES team asked all participants to answer survey questions distributed to them three times over the course of 1 calendar year. In what follows we describe in detail the participants, the data collected, and the analytical techniques employed.

Participants

As a pilot project, the project team recruited a purposive sample of 29 participants from 27 different universities from across the United States of America (see Table S1 in the online supplemental materials for details). These institutions were quite diverse—ranging from a community college offering mostly associates degrees, to doctoral universities with very high research activity. The GeoDES team recruited people through listservs of the major professional organizations in the geosciences. We received many more applicants than we had spots for, so gave preference to those who we considered to be critical voices in their department—at the minimum, those who had already earned tenure. We also gave preference for those who held or were holding significant leadership/administrative responsibilities, and expressed in their personal statement a desire to learn more about diversity, equity, and inclusion. In all, 10 participants were associate professors with tenure, and of these, only one reported not having held administrative/leadership roles in the past 3 years. There were 15 full professors with tenure in our sample, and of these, five reported not having held administrative/leadership roles within the last 3 years. One participant reported holding a rank of "distinguished professor" and also chaired the department. Finally, two participants reported "other" for their academic rank—one in which academic rank was not used at the institution (but this person was the department chair), and another reporting being a senior associate scientist and director of diversity and inclusion.

Instruments and Procedure

We administered Likert-type surveys using the online survey tool Qualtrics. See Appendix for the relevant items used for this study. Participants completed a 38-question preintervention survey (T0) before participating in the 3-day workshop. We assessed participants' self- and collective efficacy regarding diversity, equity, and inclusion. We also included the following demographic variables: (a) gender identity, (b) racial/ethnic identity, (c) academic rank held, (d) administrative positions held. Other variables not pertinent to this particular study included value beliefs about diversity, equity, and inclusion (e.g., interest value), as well as participants' beliefs about the degree to which the simulations and journal clubs were: (a) autonomy-supporting; (b) immersive; (c) interesting; (d) pressure-inducing; and (e) useful.

Participants also completed a postintervention survey 1 year after the workshop (T3; November, 2018) after completing all three simulations and journal clubs. This 42-question survey included the same items as the presurvey along with checkpoint survey items to assess participants' beliefs regarding the third simulation and third journal club meeting.

Self-efficacy. Participants' confidence in being able to identify prejudicial behaviors and confront colleagues about these behaviors was assessed pre- and postintervention using a 7-item instrument created by the research team [α = .84 (pre); 0.82 (Checkpoint 2); 0.83 (post)]. The tasks that we directed participants' attention toward with these items centered on identifying microaggressions and explicitly prejudiced behaviors, as well as confronting different types of people (nonadministrator colleagues vs. administrators). We structured these items so that we could cover a range of difficulty levels (i.e., identifying microaggressions is more challenging than identifying explicitly prejudiced behaviors), as recommended by Bandura (2006) in his guide for constructing self-efficacy scales.

Collective efficacy. Achieving certain outcomes regarding diversity, equity, and inclusion often requires collective effort among multiple people within a unit. For this reason, we assessed participants' confidence in working together with their department to create an inclusive departmental culture. Collective efficacy was assessed pre- and postintervention using a 7-item instrument created by the research team [α = .89 (pre); 0.92 (Checkpoint 2); 0.89 (post)]. The tasks that we targeted in this scale centered on "working together as a whole" with one's department because Bandura (2006) in his guide to constructing self- and collective efficacy instruments noted that an "aggregated holistic index is most suitable for performance outcomes achievable only by adept teamwork" (p. 318). Because judgment of efficacy in one's department's ability to effect change regarding diversity, equity, and inclusion is a socially embedded group endeavor rather than an individualistic one, we asked participants to rate their confidence in their department's ability to collectively accomplish certain tasks.

Analyses

Our data analytic plan was informed by the nature of the sample. Assumptions underpinning traditional null hypothesis significance testing (NHST, aka "frequentist") inferential methods tend to rely on random sampling and sufficient sample sizes to invoke the central limit theorem (van de Schoot et al., 2014). As our sample was neither random nor sufficiently large, we instead opted for a Bayesian analytic approach, which assumes that data are constant and treats estimated effects as random (Gelman et al., 2013). In the Bayesian framework, values of interest are estimated via Monte Carlo simulation and are thus not constrained by degrees of freedom concerns that limit the number of variables in a model, in relation to the sample size (McElreath, 2016). This allowed for the use of weakly informative (regularized) prior distributions on our coefficient estimates, lessening the impact of outliers on our inferences (McElreath, 2016). Use of regularizing priors may also help to limit the potential of improperly estimating the magnitude or sign of observed effects, a common issue with frequentist point estimates used in NHST on small samples (Gelman & Carlin, 2014). Taken in combination, the ability to fit models of interest to

small samples and the use of informative priors to lessen the impact of potentially spurious high-high leverage data points make a Bayesian inferential approach an ideal candidate approach for answering our research question.

Bayesian model fitting is usually conducted in four steps: (a) specify a joint distribution of the outcomes (likelihood and priors), (b) draw from a posterior distribution (usually via Markov Chain Monte Carlo simulation), (c) evaluate model fit, and (d) analyze manipulations of predictors and visualizing the results (Muth, Oravecz, & Gabry, 2018). For RQ1 we used a paired-sample version of Kruschke's (2013) proposed Bayesian estimation of mean differences, as well as to evaluate fit and visualize results, using the BayesianFirstAid package in R (Bååth, 2014; R Core Team, 2018). For RQ2–RQ4, we fitted and tested linear multiple regression models with regularized priors on the coefficients of interest using the rstanarm package in R (Goodrich, Gabry, Ali, & Brilleman, 2018). Model fit for RQ2–RQ4 were evaluated via approximate leave-one-out cross-validation (Vehtari, Gelman, & Gabry, 2017) in the loo package in R (Vehtari, Gabry, Yao, & Gelman, 2018).

We demonstrate our hypothesized data-generating models for RQ1–RQ3 below, using self-efficacy as an example.

RQ1.

$$(T2SE_i - T0SE_i) \sim t(\mu_i, \sigma_i, \upsilon_i) \quad \text{[likelihood]}$$

$$\mu_i \sim Normal\ (M_\mu, S_\mu) \quad \text{[prior for mean]}$$

$$\sigma_i \sim Shifted\ Exponential\left(\frac{1}{29}, Shift = 1\right) \quad \text{[prior for scale]}$$

$$\upsilon_i \sim Uniform(L_\sigma, H_\sigma) \quad \text{[prior for degree-of-freedom]}$$

RQ2–RQ3.

$$T2SE_i \sim Normal\ (\mu_i, \sigma_i) \quad \text{[likelihood]}$$

$$\mu_{ij} = \alpha + \beta_1 T0SE_i + \beta_2 X_i + \ldots \quad \text{[linear model]}$$

$$\alpha \sim Normal\ (0, 10) \quad \text{[prior for intercept]}$$

$$\beta_1 \sim Normal\ (0, 2) \quad \text{[regularized prior for T0SE]}$$

$$\beta_2 \sim Normal\ (0, 2) \quad \text{[regularized prior for additional covariate]}$$

$$\sigma_i \sim Exponential\ (1) \quad \text{[prior for person-level standard deviation]}$$

We answered RQ1 by examining the mean estimated differences between preintervention (T0) collective- and self-efficacy, and values of those constructs 5 months (T2) and 1 year (T3) postintervention, as well as proportion of estimated values that were greater than 0. We answered RQ2–RQ3 by examining the median estimated coefficients of interest ($\beta_2 - \beta_n$) in relation to their respective standard deviations. Median coefficient values substantively greater than their associated standard deviations (i.e., Median $> = 2SD$) were examined in more detail.

Results

We report the findings of our Bayesian model fitting analysis by Research Question (RQ). As noted in our analytic plan above we eschew making population-level inferences. Instead, we ground our following findings in the given data. Checks on model convergence and assumptions were satisfactory. For purposes of replication and validation of the model fitting procedure, please contact the authors for copies of the data and code used.

RQ1: General Trends in Self- and Collective Efficacy Over 1 Year

Figures showing longitudinal changes in efficacy beliefs are available in the online study online supplementary material. Results reveal a relatively large growth in self-efficacy between T0 ($M = 3.45$) and T2 ($M = 4.38$), along with the notable decline at T3 ($M = 3.78$). There was a similar trend in collective efficacy, with a rise between T0 ($M = 3.14$) and T2 ($M = 3.96$) and subsequent decline in T3 ($M = 3.28$). Next, we explored whether growth in mean self-efficacy from T0 to T2 is credibly different from zero. Results reveal the following: The probability of the mean difference between T0 and T2 self-efficacy ($M = 0.88$, 95% CI [0.58, 1.2]) being greater than zero was 99.9%, reflecting an average effect size of $d = 1.30$ *SD* units. Also, the probability of the mean difference between T0 and T3 self-efficacy ($M = 0.32$, 95% CI [−0.06, 0.69]) being greater than zero was 95.2%, reflecting an average effect size of 0.46 *SD* units. AQ: 6

Moving to the results for collective efficacy, results show that there is an estimated probability of 99.9% that the mean difference between T0 and T2 collective efficacy ($M = 0.69$, 95% CI [0.30, 1.1]) is greater than zero, with an average effect size of $d = 0.75$ *SD* units. However, there is an estimated probability of only 71.6% that the mean difference between T0 and T3 collective efficacy ($M = 0.11$, 95% CI [−0.30, 0.52]) is greater than zero, with an average effect size of $d = 0.13$ *SD* units. In essence, the growth in collective efficacy from T0 to T3 is not credibly different from zero.

RQ2: Interactions Between Self-Efficacy and Collective Efficacy

Growth in self-efficacy to identify prejudices and intervene appropriately did not vary based on participants' initial beliefs about their capability to work together with their department to redress DEI issues. Results show that the estimated two-way statistical interaction between T0 self-efficacy and collective efficacy was effectively zero (Median = 0, $SD = 0.1$, $d = 0.0$ *SD* units). Figures illustrating the statistical interactions between self- and collective efficacy are available in the online study online supplementary material.

However, growth in people's collective efficacy for redressing prejudicial structures in their department did vary based on par-

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

ticipants' initial self-efficacy for identifying prejudices and intervening appropriately. In fact, results reveal a detectably large statistical interaction between initial collective efficacy and self-efficacy (Median = 0.4, *SD* = 0.1, *d* = 0.42 *SD* units), with a negative direct effect of initial self-efficacy (Median = −0.4, *SD* = 0.2, *d* = −0.42 *SD* units) and positive direct effect of initial collective efficacy (Median = 0.5, *SD* = 0.2, *d* = 0.52 *SD* units), when predicting variation in 5-month postintervention (T2) collective efficacy. Although these values are difficult to interpret directly, given the interaction effect, we note that the median estimates are uniformly larger than the standard deviations of the posterior estimates, which indicates an acceptable level of precision. We point readers to the online supplementary materials for an illustration of this relationship, noting that, for participants with high (+1SD) initial self-efficacy and low initial collective efficacy (−1SD), T2 collective efficacy scores (*M* = 2.5) would likely be lower compared to peers with high (+1SD) initial self-efficacy and high (+1SD) initial collective efficacy (*M* = 4.6), reflecting an effect size difference of *d* = 2.1 *SD* units. For faculty with low initial self-efficacy (−1SD), on the other hand, T2 collective efficacy would likely not vary regardless of their initial (T0) collective efficacy.

RQ3: Do Changes in Efficacy Beliefs Vary as a Function of Demographic Factors?

In short, yes (race/ethnicity and rank) and no (gender). Controlling for preintervention self-efficacy and other demographic factors, White faculty members, on average, reported lower self-efficacy 5 months postintervention (T2) than their non-White peers (Median = −0.6, *SD* = 0.2). The median effect size difference was approximately 1.1 *SD* units. Graphs of median effect sizes are available in the online study online supplementary material. We detected no notable controlled effects of gender and faculty status on variation in T2 self-efficacy.

Controlling for preintervention collective efficacy and other demographic factors, White faculty members, on average, reported lower collective efficacy 5 months postintervention (T2) than their non-White peers (Median = −1.1, *SD* = 0.5). The median effect size difference was approximately 1.3 *SD* units. In contrast to self-efficacy, faculty members who held a rank of Full Professor, Distinguished Professor, or Emeritus Professor reported lower collective efficacy than their Associate Professor peers (Median = −0.9, *SD* = 0.3). The median effect size difference was approximately 0.9 *SD* units. We detected no notable controlled effects of gender on variation in T2 collective efficacy.

Discussion

We started with the assumption that simply delivering content to people regarding diversity, equity, and inclusion will not change people's behaviors. Thus, we assumed that training programs should emphasize practicing and reflecting on the behaviors and skills needed to confront prejudices and prejudicial structures. But practicing these behaviors is quite complex and fraught with social risks that many people are not prepared or willing to face, especially those who do not personally experience the damaging effects of prejudice. For this reason, the intervention that we designed took advantage of technological innovations that combined human conversational intelligence with artificial intelligence to simulate highly authentic situations where our participants had to actively intervene so as to generate more light than heat. For people to behave in ways that actively promote diversity, equity, and inclusion, they have to develop the efficacy beliefs to do so (Combs & Luthans, 2007; Goldstein Hode et al., 2018). We next discuss how our findings can be interpreted in light of social–cognitive theory.

RQ1: What Trends Emerged in Efficacy Beliefs Over a 1-Year Period?

Trends in self-efficacy. As hypothesized, our participants became more confident in identifying and confronting others about microaggressions and explicitly prejudicial behaviors. This self-efficacy rose sharply between T0 (November, 2017) and T2 (April, 2018), and then retreated toward the baseline by T3 (November, 2018). Despite this retreat back toward baseline, there was still credible growth from preintervention to 1-year postintervention. As we mentioned earlier, Hoy and Spero (2005) also reported that preservice teachers became more self-efficacious about teaching by the time they finished their teacher preparation program. However, when faced with the reality of having to teach their own students and to take full responsibility of a class, these same teachers became a bit more realistic about what they could and could not do. They were, however, still more self-efficacious compared to when they first started the teacher preparation program. In the same light, compared to preintervention, our GeoDES participants reported higher self-efficacy at Time 2 (5 months after the opening workshop). But when met with the reality of having to negotiate the tricky social dynamics of their home departments, our participants may have become a bit more realistic (and therefore less confident) about what they were actually capable of accomplishing.

To begin hypothesizing possible reasons for this growth and decline of efficacy beliefs, we refer to Bandura (1997), who hypothesized that self-efficacy is built through four sources: (a) *mastery experiences*, which are the interpreted results of previous successful performances; (b) *vicarious experiences* of watching similar others (or recordings of oneself) perform the same tasks; (c) *social persuasions*, which are the verbal and nonverbal assessments that influential others provide; and (d) *physiological and affective states* such as excitement or anxiety.

Of course, because at this point we do not have data to address the mechanisms behind how participants' self-efficacy rose and fell, researchers would do well in the future to collect data regarding the sources of self-efficacy. Initial investigations would have to be exploratory because, although Bandura (1997) has outlined the four sources of self-efficacy, little is known about what these sources would look like in situations involving diversity, equity, and inclusion. For example, what would be considered a mastery experience and how might that differ from social persuasions? We imagine that, when people navigate social situations involving diversity, equity, and inclusion, it is difficult to point out objective markers of mastery. Whereas there are objective measures of mastering such tasks as calculating the area of a rectangle or driving in rush-hour traffic, mastering diversity, equity, and inclusion tasks is difficult to measure objectively. We imagine, much like in teaching, that social persuasions would be one powerful

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

way to know whether one has succeeded in such situations (Morris, Usher, & Chen, 2017).

Trends in collective efficacy. Developing confidence in one's ability to identify prejudicial behaviors and confront someone who behaves this way takes much practice. However, addressing issues of diversity, equity, and inclusion in colleges and universities requires coordinating with teams of people to create new policies and collective habits that change institutional culture (Carnes et al., 2012). This is a much heavier lift compared to building people's individual self-efficacy. Nevertheless, results showed growth in participants' beliefs in their ability to work together with others to change their department's institutional culture. This collective efficacy rose sharply from preworkshop (T0) to 5 months postworkshop (T2), but then dropped back to baseline at 1 year postworkshop (T3). In fact, by the time we surveyed participants 1 year postworkshop, their collective efficacy was no different from when they started the GeoDES project.

Collective action requires a considerable amount of coordination. We hypothesize that developing faculty members' collective efficacy to address diversity, equity, and inclusion requires much more support than what we were able to provide in the GeoDES pilot project. Such support should enable department members to develop collective efficacy through mastery experiences, vicarious experiences, and social persuasions. For example, departmental goals could include increasing the percentage of faculty of color by a specified percentage within a timeframe that the group agrees on, and then assessing the department's progress toward that goal (i.e., mastery experiences). Departments could also find other departments that have been successful in achieving diversity, equity, and inclusion goals, and learn from those departments about how they were able to overcome difficulties to achieve specific goals (i.e., vicarious experiences).

RQ2: Changes in Collective Efficacy Depend on Initial Self-Efficacy

Social–cognitive theorists see individuals as proactive agents who have some power "to shape the character of their social systems" (Bandura, 2001, p. 15). Therefore, a robust self-efficacy in being able to manage one's own life circumstances has a direct effect on one's collective efficacy. This is precisely what we found in our own data, and is supported by the results of others. Recall that GeoDES participants' self-efficacy to identify and confront prejudicial behaviors had a hand in how confident they were in working collectively with their department toward improved diversity, equity, and inclusivity. That is, for participants who were highly self-efficacious to identify and confront prejudices, higher preworkshop collective efficacy predicted higher 5-month postworkshop collective efficacy. However, for those who reported low self-efficacy, initial collective efficacy had no bearing on their collective efficacy 5 months later. This finding is in line with Bandura's (2001) observation that, "one cannot achieve an efficacious collectivity with members who approach life consumed by nagging self-doubts about their ability to succeed and their staying power in the face of difficulties" (p. 16). A starting point for mobilizing changes in departmental climate, therefore, must be to ensure that individuals within the department feel capable of effectively identifying and calling in prejudicial behaviors and also acting effectively as an ally for those whose voices and contributions are devalued.

RQ3: Race and Rank Predicted Changes in Efficacy Beliefs

Race matters. We found that, despite being the overwhelming majority in the field of geosciences, the White geoscientists who participated in GeoDES were less confident compared to faculty of color in their ability to identify and call in prejudicial behaviors. Compared to faculty of color, the White geoscientists in GeoDES were also less confident in their ability to work together with their department to create more diverse, equitable, and inclusive geoscience departments. One way to interpret this finding is that, compared to their White peers, the faculty of color who participated in GeoDES bear a greater burden in changing institutional culture—they are more confident, and therefore others lean on them more to make progress in diversity, equity, and inclusion. If the White faculty who participated in GeoDES reported being less confident in identifying and redressing prejudicial behaviors, then they are less likely to speak up when their voices are needed, and also to work together with departmental colleagues to advocate for the necessary changes that create more diverse, equitable, and inclusive institutions (Bierema, 2010; Park & Denson, 2009; Sensoy & DiAngelo, 2017).

Rank also matters, but only when considering collective efficacy. Faculty members who were higher ranking (e.g., Full or Distinguished Professors) reported lower collective efficacy than did their peers who were lower ranked (Associate Professors). This could be an indication that the higher ranked faculty members who participated in GeoDES do not feel as if they have more cachet than do their lower ranked peers.

There were no differences between men and women on trends in self- or collective efficacy. On the one hand, this could mean that the women and men in our sample bore an equal burden in changing institutional culture. However, we warn against such an interpretation because the questions that we asked on the surveys clearly dealt with issues of race rather than gender. For this reason, our participants may have been more focused on their capabilities to advocate for racial diversity, equity, and inclusion. It certainly can be the case that women (especially in a field such as geosciences) feel more confident, and thus bear more of the burden, in advocating for gender diversity, equity, and inclusion (Mershon & Walsh, 2016; Ozga & Deem, 2000).

Thinking Outside the Intervention

As key institutional decision makers attempt to make sense of our findings, and to make progress on strategies to address diversity, equity, and inclusion on their own campuses, we encourage them to be open to counterperspectives regarding our findings. That is, up to this point, we have implied that variance in efficacy beliefs is explained by the activities that were part of GeoDES. Yet, we have not mentioned factors outside of GeoDES activities. For example, when we recruited participants to GeoDES we assumed that they were, in some respects, gatekeepers—faculty who had the kind of power and influence in their department to noticeably alter the departmental climate regarding diversity, equity, and inclusion. However, some consideration should be given to

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

whether GeoDES participants were actually gatekeepers. Therefore, if our participants were not gatekeepers per se, it is likely that our participants were critical voices in the department who were committed to equity and inclusion, but possibly not at the level of gatekeeper. If this suspicion is correct, then it is also likely that the participants in our program did not have the professional cachet to move the department in a new direction regarding diversity, equity, and inclusion. This would help explain the retreat back to baseline levels in collective efficacy at the 1-year mark.

Second, the retreat in efficacy beliefs toward baseline levels could be a function of participants not fully mastering the skills we were teaching them. We assumed that, because participants were engaged in the sessions and did the follow-up activities, they had mastered the content discussed during the training. In fact, we did not administer a test to determine content mastery, especially at varying levels of difficulty. Therefore, it is not out of the question that the participants' decline in efficacy beliefs at the 1-year mark could be due to a lack of strong mastery in the skills to advance diversity, equity, and inclusion goals in their department.

Third, it is possible that participants overestimated their home department's desire to advance goals in diversity, equity, and inclusion. No doubt, being confronted with the reality that departmental colleagues are not equally interested in inclusion can be a sobering and emotional discovery. Said differently, perhaps our findings merely signal that departmental desires to advance diversity, equity, and inclusion goals are limited, and require a significant disruption to make any kind of progress. When people find themselves carrying the flag for diversity alone, they could make the decision that the battle is not worth the personal costs. Lastly, it is possible that addressing collective efficacy to shift an entire department's climate regarding diversity, equity, and inclusion involves a very different set of knowledge and skills than those that are required for developing people's self-efficacy to identify and address prejudices. Modules that address self-efficacy to identify and call-in prejudicial behaviors may have to be addressed separately from those that address collective efficacy to shift institutional desire and culture regarding diversity, equity, and inclusion.

Limitations

There are a number of limitations that we acknowledge. First, given the limits in our funding, our sample consisted of a small convenience sample of 29 geoscientists who were tenured and had demonstrated some interest in learning about diversity, equity, and inclusion to further these goals in their departments. This is one reason we decided to use a Baysian analysis (as we detailed earlier in our analytical approach). As such, we caution readers not to generalize findings to contexts that are substantively different from our sample. Furthermore, although results provide preliminary evidence regarding the long-term patterns of change in faculty members' self- and collective efficacy to address issues of diversity, equity, and inclusion, a larger and more varied sample of participants should be used to verify what we report. For example, studies could involve faculty in many different fields, from a wide range of academic ranks, and a much more racially/ethnically diverse sample.

Second, our data come entirely from self-report instruments. Although these instruments have a strong track record of use in similar research contexts, self-report data are different from how people actually behave. Ultimately, as we have noted throughout, we are interested in interventions that create changes in behavior and organizational climate. Researchers would do well in the future to explore changes in participants' behaviors over time and their perceptions of departmental climate. Nevertheless, self-efficacy is an excellent proxy for people's actions (Bandura, 1997), including in situations regarding diversity, equity, and inclusion (Combs & Luthans, 2007).

Third, because we did not collect data on the sources of people's efficacy beliefs, we are unable to make any claims regarding the mechanisms behind the changes we observed in efficacy beliefs. Did efficacy beliefs rise and fall due to the presence or absence of mastery experiences? Or, could the rise and fall of efficacy beliefs be related to the topics covered by each of the simulations? That is, were participants just more confident in advocating for a job candidate (Simulation 2), but less confident in calling in microaggressions and being an ally (Simulation 1) and advocating for structural changes to the department (Simulation 3)? Our data cannot tease these issues out, but researchers would do well in future empirical investigations to test these possibilities.

Conclusion

Despite these limitations, we are able to contribute to the small but growing literature about highly immersive technology-based professional development in teaching people to actively champion initiatives in diversity, equity, and inclusion. Over a 1-year period, faculty members' confidence in being able to identify and call in prejudicial behaviors grew sharply in the first 5 months when participants were exposed to the most intensive parts of the intervention, but then retreated a bit by 1 year when the GeoDES program was coming to a close and participants were not interacting with the GeoDES leadership team as regularly. The GeoDES program appeared to have a lasting 1-year effect on participants' self-efficacy. However, collective efficacy, which is a much heavier lift because it involves people coordinating with whole departments to effect systemic changes, retreated back to baseline after a sharp rise in the first 5 months of the program. We were also able to show that developing an individual's personal self-efficacy to respond positively to diversity, equity, and inclusion has an effect on that person's beliefs about being able to work collaboratively to achieve diversity, equity, and inclusion goals for a whole department. Our findings highlight the potential for using highly immersive mixed-reality simulations that combine artificial intelligence with human conversational intuition as a way to teach faculty specific behaviors for making institutional progress on diversity, equity, and inclusion.

References

Bååth, R., (2014). Bayesian first aid: A package that implements Bayesian alternatives to the classical*. test functions in R. *Proceedings of UseR, 2014,* 2. AQ: 7

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science, 9,* 75–78. http://dx.doi.org/10.1111/1467-8721.00064

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52,* 1–26. http://dx.doi.org/10.1146/annurev.psych.52.1.1

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age Publishing.

Bernard, R. E., & Cooperdock, E. H. G. (2018). No progress on diversity in 40 years. *Nature Geoscience, 11,* 292–295. http://dx.doi.org/10.1038/s41561-018-0116-6

Bezrukova, K., Jehn, K. A., & Spell, C. S. (2012). Reviewing diversity training: Where we have been and where we should go. *Academy of Management Learning & Education, 11,* 207–227. http://dx.doi.org/10.5465/amle.2008.0090

Bierema, L. L. (2010). Diversity education: Competencies and strategies for educators. *Advances in Developing Human Resources, 12,* 312–331. http://dx.doi.org/10.1177/1523422310375024

Carnes, M., Devine, P. G., Baier Manwell, L., Byars-Winston, A., Fine, E., Ford, C. E., . . . Sheridan, J. (2015). The effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic Medicine, 90,* 221–230. http://dx.doi.org/10.1097/ACM.0000000000000552

Carnes, M., Devine, P. G., Isaac, C., Manwell, L. B., Ford, C. E., Byars-Winston, A., . . . Sheridan, J. T. (2012). Promoting institutional change through bias literacy. *Journal of Diversity in Higher Education, 5,* 63–77. http://dx.doi.org/10.1037/a0028128

Combs, G. M., & Luthans, F. (2007). Diversity training: Analysis of the impact of self-efficacy. *Human Resource Development Quarterly, 18,* 91–120. http://dx.doi.org/10.1002/hrdq.1193

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323,* 66–69. http://dx.doi.org/10.1126/science.1167311

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48,* 1267–1278. http://dx.doi.org/10.1016/j.jesp.2012.06.003

Dobbin, F., & Kalev, A. (2018). Why doesn't diversity training work? The challenge for industry and academia. *Anthropology Now, 10,* 48–55. http://dx.doi.org/10.1080/19428200.2018.1493182

Fernández-Ballesteros, R., Diez-Nicolàs, J., Caprara, G. V., Barbaranelli, C., & Bandura, A. (2002). Determinants and structural relation of personal efficacy to collective efficacy. *Applied Psychology, 51,* 107–125. http://dx.doi.org/10.1111/1464-0597.00081

Forbes. (2011). *Global diversity and inclusion: Fostering innovation through a diverse workforce*. Retrieved from https://www.forbes.com/forbesinsights/innovation_diversity/

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T. L., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology, 72,* 133–146. http://dx.doi.org/10.1016/j.jesp.2017.04.009

Freeman, R. B., & Huang, W. (2015). Collaborating with people like me: Ethnic coauthorship within the United States. *Journal of Labor Economics, 33*(S1), S289–S318. http://dx.doi.org/10.1086/678973

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science, 9,* 641–651. http://dx.doi.org/10.1177/1745691614551642

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Boca Raton, FL: CRC Press. http://dx.doi.org/10.1201/b16018

Goffee, R., & Jones, G. (2013). Creating the best workplace on earth. *Harvard Business Review, 91,* 98–106.

Goldstein Hode, M., Behm-Morawitz, E., & Hays, A. (2018). Testing the effectiveness of an online diversity course for faculty and staff. *Journal of Diversity in Higher Education, 11,* 347–365. http://dx.doi.org/10.1037/dhe0000063

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). rstanarm: Bayesian applied regression modeling via Stan (R package version 2.17.4) [Computer software]. Retrieved from http://mc-stan.org/

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . Scott-Little, C. (2012). A course on effective teacher–child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49,* 88–123. http://dx.doi.org/10.3102/0002831211434596

Herring, C. (2009). Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review, 74,* 208–224. http://dx.doi.org/10.1177/000312240907400203

Hitlan, R. T., Cliffton, R. J., & DeSoto, M. C. (2006). Perceived exclusion in the workplace: The moderating effects of gender on work-related attitudes and psychological health. *North American Journal of Psychology, 8,* 217–236.

Hoy, A. W., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21,* 343–356. http://dx.doi.org/10.1016/j.tate.2005.01.007

Hunt, V., Layton, D., & Prince, S. (2015). *Diversity matters*. Retrieved from https://www.mckinsey.com/business-functions/organization/our-insights/why-diversity-matters

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review, 71,* 589–617. http://dx.doi.org/10.1177/000312240607100404

Knoblauch, D., & Hoy, A. W. (2008). "Maybe I can teach those kids." The influence of contextual factors on student teachers' efficacy beliefs. *Teaching and Teacher Education, 24,* 166–179. http://dx.doi.org/10.1016/j.tate.2007.05.005

Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* test. *Journal of Experimental Psychology: General, 142,* 573–603. http://dx.doi.org/10.1037/a0029146

Lent, R. W., Schmidt, J., & Schmidt, L. (2006). Collective efficacy beliefs in student work teams: Relation to self-efficacy, cohesion, and performance. *Journal of Vocational Behavior, 68,* 73–84. http://dx.doi.org/10.1016/j.jvb.2005.04.001

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan* (Vol. 122). Boca Raton, FL: CRC Press.

Mershon, C., & Walsh, D. (2016). Diversity in political science: Why it matters and how to get it. *Politics, Groups & Identities, 4,* 462–466. http://dx.doi.org/10.1080/21565503.2016.1170703

Mitchneck, B., Smith, J. L., & Latimer, M. (2016). A recipe for change: Creating a more inclusive academy. *Science, 352,* 148–149. http://dx.doi.org/10.1126/science.aad8493

Morris, D. B., Usher, E. L., & Chen, J. A. (2017). Reconceptualizing the sources of teaching self-efficacy: A critical review of emerging literature. *Educational Psychology Review, 29,* 795–833. http://dx.doi.org/10.1007/s10648-016-9378-y

Moss-Racusin, C. A., van der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2014). Social science. Scientific diversity interventions. *Science, 343,* 615–616. http://dx.doi.org/10.1126/science.1245936

Muth, C., Oravecz, Z., & Gabry, J. (2018). User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *The Quantitative Methods for Psychology, 14,* 99–119. http://dx.doi.org/10.20982/tqmp.14.2.p099

Nadal, K. L. (2011). The Racial and Ethnic Microaggressions Scale (REMS): Construction, reliability, and validity. *Journal of Counseling Psychology, 58,* 470–480. http://dx.doi.org/10.1037/a0025193

Ozga, J., & Deem, R. (2000). Carrying the burden of transformation: The experiences of women managers in U.K. higher and further education. *Discourse: Studies in the Cultural Politics of Education, 21,* 141–153. http://dx.doi.org/10.1080/713661154

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400830282

Park, J. J., & Denson, N. (2009). Attitudes and advocacy: Understanding faculty views on racial/ethnic diversity. *The Journal of Higher Education, 80,* 415–438. http://dx.doi.org/10.1080/00221546.2009.11779023

R Core Team. (2018). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Sensoy, Ö., & DiAngelo, R. (2017). "We are all for diversity, but . . .": How faculty hiring committees reproduce whiteness and practical suggestions for how they can change. *Harvard Educational Review, 87,* 557–580. http://dx.doi.org/10.17763/1943-5045-87.4.557

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85,* 842–860. http://dx.doi.org/10.1111/cdev.12169

Vehtari, A., Gabry, J., Yao, Y., & Gelman, A. (2018). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models (R package version 2.0.0) [Computer software]. Retrieved from https://mc-stan.org/loo/

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27,* 1413–1432. http://dx.doi.org/10.1007/s11222-016-9696-4

West, L. (2017, July 12). Real men might get made fun of. *New York Times*. Retrieved from https://www.nytimes.com/2017/07/12/opinion/real-men-might-get-made-fun-of.html

Williams, D. A. (2013). *Strategic diversity leadership: Activating change and transformation in higher education*. Sterling, VA: Stylus Publishing.

AQ: 8

Appendix

Survey Items for Constructs Reported in Study

Self-Efficacy for Identifying and Confronting Prejudicial Behaviors

How confident are you that you can:

- identify explicitly biased behavior of your colleagues?
- identify microaggressions (subtle, unintentional, insults) of your colleagues?
- confront colleagues and peers (not an administrator) about explicitly biased behavior that they have engaged in?
- confront colleagues and peers (not an administrator) about microaggressions that they used toward another individual?
- confront administrators about explicitly biased behavior that they have engaged in?

AQ: 9

- confront administrators about microaggressions that they have used toward another in?
- enlist the support of others to confront biased behaviors of others?

Collective Efficacy for Creating More Racially/Ethnically Inclusive Departments

Working together as a whole, how confident are you that your department can:

- successfully recruit students from diverse racial/ethnic backgrounds?
- successfully recruit staff from diverse racial/ethnic backgrounds?
- successfully recruit faculty from diverse racial/ethnic background?
- successfully recruit administrators from diverse racial/ethnic backgrounds?
- create policies that promote a racially/ethnically diverse student body?
- create policies that promote a racially/ethnically diverse faculty body?
- create an atmosphere that promotes inclusivity for all members from diverse racial or ethnic groups?

Feelings of Autonomy Regarding the Simulations and the Journal Club [Scale From 1 (*Completely False*) to 6 (*Completely True*)]

- The (interactive simulation/journal club) provided me with interesting options and choices
- I experienced a lot of freedom in the (interactive simulation/journal club).
- During the (debriefing session/journal club), I felt like I was free to talk about things that I wanted to discuss.

(*Appendix continues*)

tapraid5/dhe-dhe/dhe-dhe/dhe99920/dhe0367d20z | xppws | S=1 | 4/27/20 | 9:52 | Art: 2019-0589

Interest in the Simulations and the Journal Club [Scale From 1 (*Completely False*) to 6 (*Completely True*)]

- I enjoyed participating in the (interactive simulation/journal club).
- The (interactive simulation/journal club) was a fun activity to do.
- I would describe the (interactive simulation/journal club) as very interesting

Beliefs about the Usefulness of the Simulations and Journal Club [Scale From 1 (*Completely False*) to 6 (*Completely True*)]

- I believe that participating in the (interactive simulation/journal club) was valuable to my professional growth.
- I would be willing to participate in an (interactive simulation/journal club) like this again because this activity was useful to me.
- The (interactive simulation/journal club) taught me beneficial skills that I could use in my department.

Feelings of Immersion/Presence in the Simulations [Scale From 1 (*Completely False*) to 6 (*Completely True*)]

- When interacting in the simulation, I felt like I was actually interacting with real people.
- When interacting in the simulation, I felt like I had been transported to a real-life situation.
- The characters in the simulation seemed to represent real people I have interacted with in the past.
- The scenario that the experience simulated felt like a real-life situation that I could be faced with in the future.
- The actions of the avatars seemed like ones that I might have to confront in the future.

Perceived Competence in the Simulations [Scale From 1 (*Completely False*) to 6 (*Completely True*)]

- I thought I did pretty well in the interactive simulation.
- Compared to other participants, I thought I did pretty well in the interactive simulation.
- I am satisfied with how well I did in the interactive simulation.
- I was pretty skilled in the interactive simulation.

Feelings of Pressure/Tension in the Simulations [Scale From 1 (*Completely False*) to 6 (*Completely True*)]

- I felt very tense while I was participating in the interactive simulation.
- I was anxious while I was participating in the interactive simulation.
- I felt pressured while participating in the interactive simulation.

Received August 11, 2019
Revision received March 31, 2020
Accepted April 7, 2020 ■

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES 1

AQau—Please confirm the given-names and surnames are identified properly by the colors. = Given-Name, = Surname
The colors are for proofing purposes only. The colors will not appear online or in print.

AQ1—Author: Please confirm that running header is acceptable or provide alternate header of 50 characters max, including spaces.

AQ2—Author: Please be sure to provide the name of the department(s) with which you and your coauthors are affiliated at your respective institutes if you have not already done so. If you are affiliated with a governmental department, business, hospital, clinic, VA center, or other nonuniversity-based institute, please provide the city and U.S. state (or the city, province, and country) in which the institute is based. Departments should be listed in the author footnote only, not the byline. If you or your coauthors have changed affiliations since the article was written, please include a separate note indicating the new department/affiliation: [author's name] is now at [affiliation].

AQ3—Author: Your article contains supplemental material. If you have queries for references that are not cited in the text, these can be ignored if the references are found in the supplemental material.

AQ4—Author: Please check if correct as edited: "...simulations in which human conversational intuition combines with..."

AQ5—Author: T0, T2, and T3 spelled out on first use per APA style guidelines.

AQ6—Author: 95% Confidence intervals formatted according to APA style guidelines.

AQ7—Author: Please provide either publisher location and name, or URL for material in reference (Bååth, 2014) ... and page range.

AQ8—Author: Please provide an appropriate title for the appendix.

AQ9—Author: Lead sentence ends with "can" ... so "can" removed from the beginning of questions.