ANNUAL REVIEWS

Annual Review of Psychology

Diversity Training Goals, Limitations, and Promise: A Review of the Multidisciplinary Literature

Patricia G. Devine[1] and Tory L. Ash[2]

[1]Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA; email: pgdevine@wisc.edu

[2]Department of Psychology, Syracuse University, Syracuse, New York 13244, USA; email: tash02@syr.edu

Annu. Rev. Psychol. 2022. 73:403–29

First published as a Review in Advance on July 19, 2021

The *Annual Review of Psychology* is online at psych.annualreviews.org

https://doi.org/10.1146/annurev-psych-060221-122215

Keywords

diversity training, cultural competence, diversity education, prejudice reduction interventions, unconscious bias training, bias literacy

Abstract

In this review, we utilize a narrative approach to synthesize the multidisciplinary literature on diversity training. In examining hundreds of articles on the topic, we discovered that the literature is amorphous and complex and does not allow us to reach decisive conclusions regarding best practices in diversity training. We note that scholars of diversity training, when testing the efficacy of their approaches, too often use proxy measures for success that are far removed from the types of consequential outcomes that reflect the purported goals of such trainings. We suggest that the enthusiasm for, and monetary investment in, diversity training has outpaced the available evidence that such programs are effective in achieving their goals. We recommend that researchers and practitioners work together for future investigations to propel the science of diversity training forward. We conclude with a roadmap for how to create a more rigorous and relevant science of diversity training.

Contents

INTRODUCTION

Public discourse and popular media are flooded with stories of companies implementing diversity training (DT) in response to highly publicized, and often reputation damaging, instances of bias. In a particularly salient case, Starbucks closed all 175,000 stores to host a four-hour antibias training following the controversial arrest of two Black patrons purportedly loitering while waiting for an associate to arrive (Stewart 2018). In response to the outrage following the expulsion of Black passengers for allegedly laughing too loudly on the Napa Valley Train, the CEO of the company publicly promised to provide DT for all his employees (Bhattacharjee 2016). In another example, Delta Air Lines offered unconscious bias training for all 23,000 flight attendants after a Black physician's credentials were questioned when she attempted to provide emergency medical care to a fellow passenger (Crespo 2018). Although the public is all too familiar with promises of reform via employee DT, much less attention is paid to the content, objectives, and effectiveness of DT. Specifically, what are the goals of DT? What should be included in DT? How would an interested consumer recognize an effective DT program? And is DT effective in reducing bias, or is it rife with empty promises?

Many scholars and laypeople alike argue that DT may be effective across a variety of contexts for reducing intergroup anxiety, preventing discrimination, and ultimately, promoting social justice. We find the hunger for knowledge regarding what practitioners can do to create more inclusive environments encouraging, as motivation to address bias is the necessary first step to achieving greater equity (Devine 1989). However, the well-intentioned, yet uninformed, consumer may quickly become overwhelmed by the breadth of DT programs currently available. DT programs go by many names and range from diversity and inclusion certification programs at accredited

universities to bias training via online modules and consultation services from diversity, equity, and inclusion experts. Although many programs boast endorsements from well-known companies that vouch for the efficacy of their services, their websites provide little evidence supporting the effectiveness of their programs. Despite stylish web pages featuring photos of diverse work teams, lofty promises, and persuasive customer testimonials, there is a lack of information about the particular content, techniques, and evidentiary basis underlying the application of each training. Moreover, diligent browsers are often frustrated in their efforts because many websites prohibit prospective clients from gaining more information without signing up for a listserv, consultation, or free trial.

Despite the abundance of DT programs available to purchase by the public, the practice of offering DT has gotten too far ahead of the evidence suggesting they are helpful (e.g., Green & Hagiwara 2020, Moss-Racusin et al. 2014, Paluck et al. 2020). Furthermore, some scholars explicitly question the ethics of implementing such trainings without evidence of their efficacy (Paluck 2012). Others have sounded the alarm that such trainings may even be counterproductive and may be associated with a decrease in the representation of employees from historically marginalized groups (Dobbin & Kalev 2016, Dover et al. 2020).

This review summarizes the goals, content, and efficacy of DT across a variety of disciplines and settings. In light of the boom in DT and in response to calls for a more rigorous evaluation of the efficacy of DT programs (Paluck 2012, Paluck & Green 2009), this review focuses on the extent to which the science of DT has gained traction in establishing the efficacy of DT programs. And if not, where do we go from here?

METHOD

Given that others have noted that "diversity training" can be considered a catch-all term (e.g., Paluck 2006), we cast a wide net in performing our literature search on DT. Articles included in our review evaluated DT programs targeted to address outcomes relevant to institutionalized settings. All of the studies reviewed share an emphasis on relevant samples (i.e., nurses, teachers, employees), field settings (i.e., classroom, workplace, professional conference), and training programs (rather than brief lab-based manipulations). This review is distinct in highlighting DT, specifically, and it departs from previous reviews that examine the effects of contact (Pettigrew & Tropp 2006) or the broad array of prejudice reduction manipulations designed to enhance intergroup relations (Paluck et al. 2020).

We used a variety of search terms and did not restrict our search to any particular field or set of journals. Search terms included: diversity training, bias/prejudice reduction interventions, antibias training, diversity education, cultural competence, bias literacy, multicultural education, ethnic studies, implicit/unconscious bias training, and racial sensitivity training. We limited our search to articles that were peer-reviewed, had adult samples, and were published during or after the year 2000. Although DT for children and adolescents is a growing topic of inquiry, this body of literature involves considerations (e.g., the developmental appropriateness of the program's content) that fall outside the scope of this review. We restricted our database to articles published after the year 2000 for two reasons. First, that year largely marks the beginning of the big business boom of DT as a for-profit and pervasive industry (Paluck 2006). Second, comprehensive reviews of the DT literature prior to 2000 already exist (e.g., Bezrukova et al. 2016, Paluck & Green 2009). Our goal was to evaluate the extent to which the more recent science of DT has progressed to the point of offering clear guidelines regarding best DT practices.

Our literature search began in June of 2019 and continued until the end of 2020. In total, we collected 250 articles, which were then coded across 35 different criteria. The majority of the coding was conducted by the second author; all other coding was conducted by trained

research assistants and checked by the second author. To obtain interrater reliability, two coders independently reviewed all of the articles and coded for 6 of the 35 variables, for a total of approximately 15% of the data. These 6 variables correspond to the findings reported throughout our review; interrater reliability was satisfactory (97.60% agreement).

Variables of interest were selected as being likely important for evaluating the effectiveness of DT based on prior literature reviews and meta-analyses (e.g., Bezrukova et al. 2016, Paluck & Green 2009, Pettigrew & Tropp 2006). We distinguished articles based on their setting, purpose, kind of training, and duration. To account for the scientific rigor of articles, we coded for the research design utilized, the sample selection and size, whether outcomes were self-reported or behavioral, and whether assessments were immediate or delayed. Throughout, we highlight the variables that are most germane for our review; however, readers interested in learning more about the other variables can do so on our page on the Open Science Framework website (**https://osf.io/p7sxr/**).

Our review includes studies that were conducted in one of three settings—organizational, human services, and education—each with its own definition of DT and specific goals that the DT is meant to address. Studies conducted in organizational settings concerned diversity initiatives for employees in workplace settings. Articles within the subfield of human services discussed training for service providers (e.g., doctors, mental health professionals, and teachers) to promote equitable care. And studies positioned in educational settings evaluated the efficacy of diversity-related curricula directed at a general student audience.

Each of these subfields has an extensive DT literature, and evaluating them separately allows for an analysis of the unique strengths and shortcomings of the research in each context. In organizing our review around these subfields, we depart from prior meta-analyses that include DT but do not make such distinctions (e.g., Bezrukova et al. 2016, Paluck & Green 2009, Paluck et al. 2020). Evaluating whether DT is effective requires considering the specific goals that motivate the implementation of particular trainings. Trainings implemented to increase minority representation in a workplace, for example, have different objectives, targeted outcomes, and content compared to a training aimed at reducing patient treatment discrepancies in health care settings. As such, making direct comparisons across subfields is challenging, and inferences regarding DT made in one discipline may or may not generalize to another subfield.

Due to the disparate methodologies and wide-ranging practices encompassed by the cross-disciplinary term "diversity training," we used a narrative approach in summarizing the literature. Within each discipline we identify the goals and approach of DT for that field, the most common methods used, the outcomes assessed, and the state of the evidence regarding the efficacy of DT. We then offer a critique of the work and some field-specific recommendations for advancing the science of DT. We conclude each section with a table summarizing the work done in that particular field and our field-specific recommendations.

DIVERSITY TRAINING FOR EMPLOYEES IN ORGANIZATIONAL SETTINGS

We begin our summary of DT with studies positioned in workplace or organizational settings. Within the United States, employee DT was born in response to the advent of affirmative action policies implemented in US workplaces following the civil rights movement of the 1960s and 1970s. During this time, DT was simply used to inform employees of antidiscrimination laws and to assimilate women and people of color into workplace culture. Today, many motivations likely underlie companies' utilization of DT, such as the promotion of a diverse workforce, the provision of effective communication with a diverse customer base, the avoidance of workplace

discrimination, and the cultivation of creative problem solving. Irrespective of motivation, as demographics continue to shift, corporations are tasked with creating increasingly multicultural, multiracial, and multigendered workplace communities. As a result, DT has become a big and booming industry. Undeniably, DT sells, and it sells well; by one estimate, companies invest $8 billion in DT each year (Lipman 2018). Currently, more than half of mid-sized and large US companies offer some form of DT (Dobbin & Kalev 2016). What is unknown, at this point, is whether the returns, in terms of benefits, warrant the huge investments in DT.

Goals and Approach

The goals for organizational DT include the "full integration of members of minority social categories into the social, structural, and power relationships of an organization or institution" (Brewer et al. 1999, p. 337). These goals encompass the recruitment and retention of employees from underrepresented backgrounds as well as increased group cohesion, creativity, and equity within a given workplace. Stated simply, organizational DT has the overarching goal of fostering an inclusive company climate (Bezrukova et al. 2016). Therefore, our review of the literature evaluated questions such as, Does DT lead to increased feelings of belonging among members of historically marginalized groups? Does representation of members of historically marginalized groups improve following DT, and is this increase maintained over time? Do employees from both historically advantaged and disadvantaged groups who undergo DT report more inclusive work climates, compared to employees from organizations that do not offer DT?

Articles are included in this section if they discuss topics or use samples characteristic of organizational settings. For example, Combs & Luthans (2007) studied participants from a government agency, insurance company, and manufacturing firm. Others investigated government contract trainees (Rehg et al. 2012), managers within a government agency (Sanchez & Medkik 2004), and even taxicab drivers (Reynolds 2010). Many studies recruited business graduate students (e.g., Bush & Ingram 2001, Sanchez-Burks et al. 2007), hospitality students (Madera et al. 2011), and undergraduate students either enrolled in a workplace diversity course (Hostager & De Meuse 2008) or engaged in a professional setting as research (Roberson et al. 2009) or teaching (Roberson et al. 2001) assistants. A sizeable portion of the articles (17.02%) examined the impact of trainings targeting gender bias in science, technology, engineering, and math (STEM) organizations and departments (e.g., Hennes et al. 2018, Moss-Racusin et al. 2018).

Many trainings pertained to the general promotion and inclusion of marginalized groups. Others, however, were specific about the group targeted, such as women (e.g., Chang et al. 2019, Jackson et al. 2014), older individuals (Reynolds 2010), English language learners (Madera et al. 2011), and individuals with disabilities (Phillips et al. 2016).

DT within organizational settings is most commonly delivered in a lecture-based format by an outside consultant (Paluck 2006). Throughout the presentation, trainers often discuss the definition, benefits, and potential challenges of workplace diversity. The presentation is typically followed by group activities, such as reviewing cases of work-based prejudice (Sanchez & Medkik 2004), simulating common disabilities associated with aging (Reynolds 2010), and determining whether different scenarios constitute workplace discrimination (Preusser et al. 2011).

As found by prior reviews, the selection of particular DT strategies appears to be most often motivated by personal preference or intuition about what trainers believe would be effective rather than by a specific theoretical approach or empirical evidence (Cox & Devine 2019, Pendry et al. 2007). Many studies from organizational settings did not include information explaining the content of the training (e.g., Holladay & Quiñones 2008) or justifying the use of the strategies employed (e.g., Sanchez & Medkik 2004).

Research Designs and Assessment of Outcomes

Of the 47 articles that discussed DT in organizational settings, 15 articles were correlational, theoretical, or qualitative; 32 studies delivered and quantitatively evaluated a training. Researchers most often utilized single-group repeated measures designs (i.e., pre–post; 43.75%), others used group designs with random assignment (i.e., experimental; 37.50%), and a few utilized group designs without random assignment (i.e., quasi-experimental; 18.75%).[1]

Most studies (62.50%) assessed trainees' cognitive and affective responses to the DT as their primary outcome of interest.[2] Specifically, many studies' primary outcome was employees' self-reported learning or recalled knowledge of the material presented within the training, such as knowledge of how stereotypes may influence one's judgments in professional settings (Roberson et al. 2009), what constitutes bias (Hennes et al. 2018), and knowledge of cultural differences (Rehg et al. 2012). Several studies also examined trainees' perceptions of the training itself (e.g., liking, interest) through a program evaluation survey (e.g., Reynolds 2010, Sanchez & Medkik 2004). Other studies examined participants' attitudes following DT, such as supportive attitudes toward women in the workplace (e.g., Chang et al. 2019), attitudes toward LGBTQ+ employees (e.g., Hood et al. 2001), and attitudes toward non-English speakers (Madera et al. 2011).

Although most studies focused on trainees' self-reported outcomes, some studies (28.13%) did not. Instead, these studies examined how the DT affected participants' responses to hypothetical workplace diversity incidents (Roberson et al. 2009), supervisors' ratings of trainees' interpersonal skills (Sanchez & Medkik 2004), and creative problem solving within nationally diverse teams (Homan et al. 2015). In a methodologically rigorous study, Chang and colleagues (2019) assessed the impact of the training on the number of female employees nominated for excellent performance in an ostensibly unrelated workplace initiative.

State of the Evidence

Although many trainings demonstrated favorable post-intervention effects with respect to employees' self-reported cognitive, affective, and skill-based outcomes (Kalinonski et al. 2013), other studies demonstrated more complicated patterns of results. For example, in a quasi-experimental field study, Sanchez & Medkik (2004) found that diversity awareness training actually led to an increase in managers' unfriendly treatment toward non-White employees, as rated by a coworker specifically assigned to monitor the behavior of each participant. Based on post-intervention interviews, the authors concluded that adverse outcomes arose out of resentment because the trainees believed they had been referred to the mandatory DT following complaints of biased behavior.

Studies that incorporated delayed measures (40.63%) found conflicting evidence that immediate effects translated into enduring changes. For example, Chang and colleagues (2019) investigated the effects of an online DT and found that although some participants reported more positive attitudes toward women immediately after the intervention, there was limited evidence to support the training's efficacy in delayed behavioral measures collected 3 weeks later. Hill & Augoustinos (2001) studied a program aimed at reducing prejudice toward Aboriginal

[1] Studies were coded once for design, and the most rigorous design feature was recorded. For example, although many studies with random assignment to condition included pre- and post-assessments, these were coded as experimental.

[2] Although most studies included multiple outcomes, the studies were coded for their primary outcome, as specified in the abstract, which usually coincided with the variable that demonstrated effects (Paluck et al. 2020). However, in the case of studies that included a behavioral or implicit (i.e., not self-reported) outcome, this was prioritized as the primary outcome of interest.

Australians among employees in a large Australian-based public service organization. Although negative stereotypes and prejudice decreased immediately following the program, these changes did not persist when evaluated 3 months later. Adding to the mixed nature of the findings, Combs & Luthans (2007) found that 1 year after the training, participants who had received DT valued diversity more compared to employees in the control condition. Given that the differences across these studies are numerous (i.e., the particular organizational setting of the training, the content of the DT, the outcomes examined, etc.), it is unclear what underlies the differences observed regarding the long-term effects of DT.

Contradictory findings concerning the impact of DT in organizational settings may highlight the importance of contextual factors that can moderate a program's effects, such as whether training is mandatory or optional. Research has indicated that compulsory DT can often result in backlash (e.g., Legault et al. 2011, Sanchez & Medkik 2004) and may actually lead to less diversity in the workforce (Dobbin & Kalev 2016). However, voluntary training may only benefit participants who already appreciate diversity at the onset (Kulik et al. 2007). Although the mandatory versus voluntary training debate is beyond the scope of this review, recent research has provided a potential avenue of resolution. Rather than assigning all members of an organization to attend DT, organizations may be better served by equipping socially connected and highly respected individuals with the tools and motivation to inform and persuade other members of a social network to promote greater equity (Forscher 2017, Paluck et al. 2016).

In our review of the literature, measurements of systemic bias—such as minority representation, prevalence of workplace discrimination, and the promotion rates of historically marginalized employees—were largely absent. Of the articles that adapted a systems-level lens, Waight & Madera (2011) found that in a survey of hospitality employees, offering workplace DT was positively related to job satisfaction, was negatively associated with perceived workplace discrimination, and reduced turnover intentions, but only for employees from historically marginalized groups. In a rigorous cluster-randomized, controlled trial, researchers at the University of Wisconsin–Madison implemented a theoretically and empirically motivated gender bias habit-breaking intervention within STEM departments and evaluated the effects not only on self-reported outcomes (e.g., awareness of bias, self-efficacy to address bias) but also on departmental climate, as assessed in an unrelated annual survey of workplace climate conducted within the university. Faculty in intervention departments reported better fit, felt that their scholarship was more valued by colleagues, and felt more comfortable raising family obligations than did faculty in control departments (Carnes et al. 2015). In an evaluation of the training 2 years later, Devine and colleagues (2017) found that intervention departments demonstrated increased hiring of female faculty compared to control departments.

Taken as a whole, our review of the literature on DT reveals that, in light of the overarching goals of DT in these settings, the evidence regarding the efficacy of DT is for the most part wanting. The lack of systemic and rigorous research investigating company-wide DT, combined with the mixed nature of evidence regarding the efficacy of the programs, prevents us from drawing clear conclusions regarding best practices for organizational DT.

Limitations and Recommendations

Though the evidence amassed to date is limited, it provides some clues as to how to move forward to deliver better investigations of DT within organizations. In evaluating diversity initiatives within organizational settings, researchers must focus on the stated goals of the programs. The common indicators of success seem to be the completion of the program and its favorable evaluation by the trainees, rather than clear progress toward the program's targeted goals. Given the significance of workplace DT for fostering inclusion and comfort in the face of a diversifying

workforce, it is important to evaluate if these goals are being met using scientific methods that allow for testing these hypotheses. Research designed to test the intended objectives of DT should employ large-scale, longitudinal, and contextually relevant methodology as well as objective indicators of success, such as the representation, retention, and advancement of employees from historically marginalized groups (Moss-Racusin et al. 2014, Paluck 2006). In contrast, our review revealed an overreliance on immediate, self-reported, and individual-level measures that cannot speak to the systems-level goals of DT.

Our review brought into sharp relief the distinction between two types of measures—individual- and systems-level outcomes—that are used to evaluate the extent to which DT brings about change. Although the overarching goals of DT are to create systemic changes (i.e., retention of historically marginalized employees, improved perceptions of workplace climate, decreased frequency of workplace discrimination), researchers most often assess outcomes at the level of the individual (e.g., self-reported measures of knowledge, liking of the program, and attitude), which are often taken as evidence of the training's effectiveness under the assumption that individual-level changes will translate into systems-level changes. However, individual-level, self-reported cognitive and affective outcomes are, at best, indirect indicators of the intended systems-level changes.

This measurement problem is not unique to DT research. It is well documented, for example, in clinical research when researchers examine the impact of an intervention or treatment on an ultimate outcome by assessing surrogate measures, which are theoretically related outcomes that are often easier, faster, or less costly to measure (VanderWeele 2013). It is likely the ease with which self-reported attitudes and diversity-related knowledge are assessed that led to their continued use as a surrogate outcome for inclusive workplace environments. Although the relationship between individual attitudes and behaviors seems intuitive, the literature on the correspondence between attitudes and behaviors reveals a more complicated relationship (LaPiere 1934, Wicker 1969). A burgeoning body of literature exposes the substantial disconnect between individuals' self-reported prejudice-related attitudes and their observed discriminatory behaviors (Forscher et al. 2019, Paluck et al. 2020).

It is the responsibility of researchers within organizational DT who continue to employ individual-level measures to establish the validity of these measures as surrogate indicators of the ultimate and systems-level outcomes of interest. Evaluating the efficacy of DT training programs requires either (*a*) demonstrating that individual-level outcomes are directly related to the systems-level changes or (*b*) assessing system-level outcomes over time to reveal if the training is truly effective in creating an enduring improvement in the experiences of historically marginalized individuals within an organization. In so doing, researchers can conduct more relevant investigations of DT and better justify the use of individual attitudes and knowledge as a reasonable surrogate measure when evaluating whether the goals of a particular DT are met within an organization.

In support of emphasizing systems-level change, converging evidence suggests that individual DT was more impactful on surrogate outcomes when delivered alongside larger workplace diversity initiatives (Bezrukova et al. 2016) or when openly supported by upper-level management (Rynes & Rosen 1995). This finding is consistent with other researchers' observations: The effectiveness of DT is limited when company policy does not reflect the concerns of people from traditionally underrepresented groups (Dobbin & Kalev 2016, Pendry et al. 2007). When company policies appreciate and advocate for historically marginalized employees, it signals the company's values, scaffolds the creation of prosocial norms, and communicates authorities' explicit commitment to creating an inclusive company climate. As such, any effective DT should be implemented in tandem with leadership endorsement of diversity initiatives that promote employees

Table 1 Summary of the literature and our recommendations for DT in organizational settings

DT in organizational settings	
Goals and approach	■ The goal is to promote inclusive workplace climates and to increase the recruitment, retention, and perceived belonging of employees from historically marginalized groups. ■ The training is primarily composed of content-based instruction (lectures, pamphlets, etc.) with some interactive components (discussions, exercises, role playing, etc.).
Research designs and assessment of outcomes	■ Some studies (37.50%) used experimental designs, a few (18.75%) used quasi-experimental designs, and others (43.75%) used pre–post designs. ■ Some studies (40.63%) used delayed follow-up measures. ■ Most studies (62.50%) assessed employees' individual-level cognitive and affective responses to the DT as the primary outcome of interest. ■ Few studies (28.13%) examined outcomes that did not rely on self-report.
State of the evidence	■ There is mixed evidence of the impact of DT at immediate posttest assessment. ■ There is conflicting and limited evidence regarding the long-term effectiveness of organizational DT. ■ Evidence is inconclusive regarding whether diversity-related programming promotes inclusive workplace environments or increases the recruitment, retention, and perceived belonging among employees from historically marginalized groups.
Limitations	■ Justification for the training methods employed is lacking. ■ There is a mismatch between the goals of the training and the outcomes used to evaluate its efficacy (i.e., overreliance on surrogate or proxy measures). ■ Many factors likely contribute to mixed findings, such as whether the DT is made mandatory by management and whether DT is delivered alongside other diversity-related initiatives.
Recommendations	■ Rather than making DT voluntary or mandatory, consider targeting socially connected individuals within an organization. ■ Utilize behavioral and systems-level outcomes to ensure the goals of training are being met. ■ Monitor the hiring, retention, and perceived belonging of employees from historically marginalized groups over time. ■ Embed training in larger workplace diversity initiatives with support from the upper management.

Abbreviation: DT, diversity training.

from diverse backgrounds rather than delivered in a "one-and-done" approach (see **Table 1** for a summary of the literature and our recommendations for DT in organizational settings).

DIVERSITY TRAINING FOR HUMAN SERVICE PROVIDERS

Members of historically marginalized groups experience lower quality of care and are less likely to receive routine and preventative treatments; they also experience greater difficulties accessing adequate mental health services. As a result, members of marginalized groups face higher rates of morbidity and mortality than nonminority individuals do (Carratala & Maxwell 2020).

To achieve equitable care, many recommend training human service providers to deliver culturally competent care. In the United States, the Office of Minority Health has developed the National Standards on Culturally and Linguistically Appropriate Services, which continue to be adopted throughout the country. According to these guidelines, culturally competent care takes into account clients' cultural beliefs, health literacy, and communication needs to provide respectful, accessible, and equitable services. These criteria were developed with the ultimate goals of advancing equity within human services, improving the quality of interactions with clients, and eliminating extant disparities within health-related settings.

Goals and Approach

During cultural competence trainings, trainees are provided with skills and knowledge presumed to bridge cultural divides and facilitate effective intervention despite a cultural mismatch between clients and providers. DT is proposed to be a vehicle by which to improve the experiences of marginalized clients within human services and reduce inequities in health-related outcomes. Given these goals, as we reviewed the literature we asked, Does cultural competence training lead to improved quality of care in services received by historically marginalized clients? Do culturally competent providers achieve more equitable client outcomes compared to providers without cultural competency training?

Our search for relevant literature yielded 142 articles. The majority of the studies examined the development of cultural competence among medical, health service psychology, social work, and nursing students. Others examined outcomes for health care professionals, including nurses (e.g., Berlin et al. 2010, Brathwaite & Majumdar 2006), hospice staff (Schim et al. 2006), practitioners who specialize in sickle cell disease (Thomas & Cohn 2006), and individuals who work specifically with culturally and linguistically diverse communities (Henderson et al. 2011). Additionally, several studies examined cultural competence training for mental health and wellness professionals, such as counseling graduate students (e.g., Kagnici 2014), alcohol and drug counselors (Luger 2011), clinical managers (Abernethy 2005), and occupational therapists (Leyva et al. 2014).

We also included articles (11.97%) that involved trainings for teachers and preservice teachers. Although teaching does not fit neatly with the type of human service provision described previously, the type of DT most often conduced with teachers has goals consistent with cultural competence training. Namely, DT for educators is aimed at cultivating teachers' cultural competencies to improve interactions with historically marginalized students and mitigate widespread disparities within education.

Cultural competence training aims to increase providers' knowledge of culturally based beliefs that may influence clients' experiences with human services. As one example, providers are taught that Hispanic communities may endorse *fatalismo*, or a belief that health and illness are a product of destiny rather than the object of proactive control, and *familismo*, or an emphasis on the importance of family input in forming treatment decisions (Flores 2000). Other kinds of DT include antiracism training for child welfare employees (Johnson et al. 2009), a weight stigma reduction intervention for clinical psychology trainees (Brochu 2020), and DT concerning equitable care for members of the LGBTQ+ community in a senior care facility (Holman et al. 2020).

Although the majority of studies implemented cultural competence training for human service providers and trainees, the methods employed varied. Many studies (e.g., Carter et al. 2006, LoboPrabhu et al. 2000) had the trainees role-play clinicians treating patients from different cultural groups, and the trainers provided feedback on the trainees' cultural sensitivity. Dogra (2001) assigned undergraduate premedical students a disability, such as blindness or hearing impairment, and encouraged students to contemplate the positive and negative aspects of having the disability. Some studies emphasized the importance of navigating language barriers in communicating with linguistically diverse patients (e.g., Henderson et al. 2011, Xu et al. 2010), and one even provided health care workers with foreign language courses (Mazor et al. 2002). One study promoted lessons and structured interactions relevant to caring for refugee families (Griswold et al. 2006). Other trainings relied on more participatory learning, in the form of home visits (Juarez et al. 2006), cultural immersion (Diaz-Lazaro & Cohen 2001), and community-based clinical practicum placements (Amerson 2010).

DT for human service providers emphasizes the importance of increasing both the knowledge and the skills relevant to culturally competent care. Rather than simply teaching trainees about

widespread inequities, DT often teaches participants tangible skills that can be implemented to provide more equitable services. For example, one training focused on the importance of communication skills aimed at building trust in patients with sickle cell disease who feel negatively labeled by the health care system (Thomas & Cohn 2006). Similarly, Hughes & Hood (2007) provided nursing students with skills in interview-based cultural assessment to better identify the needs of, and develop a plan of care for, culturally diverse patients. Schim and colleagues (2006) accentuated the importance of hospice workers' communication skills by scaffolding active listening and use of accessible language in patient interactions. Luger (2011) taught alcohol and drug counselors to identify cultural factors related to mental health stigma and risk in patient assessment and intervention.

Research Designs and Assessment of Outcomes

In our sample of articles concerning DT for human service providers, 98 studies implemented and quantitatively evaluated trainings delivered by researchers. Of these 98 studies, most studies evaluated outcomes pre- and post-intervention (66.33%), others were quasi-experimental (20.41%), and a few were experimental (13.27%). Although the majority of studies relied only on measures collected immediately posttest, a few studies in our sample included delayed assessments of outcomes (23.47%).

Cultural competence trainings are designed to improve outcomes for clients. Nevertheless, the vast majority of studies (85.71%) utilized human service providers' self-reported ratings as the primary outcome of interest. The most common outcome assessed was providers' self-reported cultural competence through the use of established surveys. Other studies assessed trainees' confidence in interacting with culturally diverse patients, such as health care providers' self-efficacy in communicating with stigmatized patients (e.g., Thomas & Cohn 2006) or students' transcultural self-efficacy (e.g., Amerson 2010). A few studies examined changes in implicit (e.g., Castillo et al. 2007) and explicit (e.g., Crandall et al. 2003) attitudes toward marginalized groups following a cultural competence intervention.

The emphasis on measuring providers' cultural competence following DT reflects the assumption that these self-reported outcomes translate into actions that will lead to improved outcomes for clients from historically marginalized groups. However, only a minority of studies (13.27%) tested this assumption by evaluating the impact of DT using behavior-based and systems-level outcomes. As one example, during a 4-month follow up, Prescott-Clements and colleagues (2013) evaluated the impact of their intervention on trainees' responses to standardized patient scenarios in which actors played patients making inappropriate remarks, experiencing communication difficulties, or having religious concerns about a recommended treatment. Other studies examined patient outcomes directly, such as patient satisfaction (Mazor et al. 2002), patient utilization of health and social services (Majumdar et al. 2004), and patient health outcomes (Thom et al. 2006).

State of the Evidence

Across the majority of studies surveyed, with some exceptions (e.g., Beagan 2003), the evidence suggests that cultural competence training was associated with increases in human service providers' self-reported cultural competence (e.g., Beach et al. 2005, Renzaho et al. 2013). Results showcasing the efficacy of cultural competence training in promoting knowledge of cross-cultural client care are promising. However, there is less information available about how these improvements translate into provider behaviors that are likely to influence the experiences of clients from historically marginalized groups. The evidence suggests that the cultural competence of human

service providers was both associated with (Castro & Ruiz 2009, Majumdar et al. 2004, Weech-Maldonado et al. 2012) and unrelated to (Thom et al. 2006) positive patient outcomes, such as patient satisfaction, utilization of treatment resources, and patient trust. Given the contradictory findings and the paucity of research on client outcomes, future research should consider client perspectives as a primary outcome when evaluating the impact of trainings for human service providers (Lie et al. 2011, Renzaho et al. 2013).

Similarly, studies that evaluated training for preservice teachers found that DT elevated self-reported cultural competence (Rogers-Sirin & Sirin 2009), decreased stereotypic attitudes (Amatea et al. 2012), and led to more positive attitudes toward diversity (Middleton 2002). Few studies examined outcomes that extended beyond preservice teachers' self-reported attitudes and beliefs; however, those that did found evidence of the training's impact in participants' responses to videotaped school-based ethical dilemmas (Rogers-Sirin & Sirin 2009) and teaching case conceptualization (Amatea et al. 2012). No studies in our review examined student outcomes in validating the efficacy of a particular DT for preservice or current teachers.

As was true of studies examining DT within organizational settings, researchers evaluating DT for human service providers were overly reliant on surrogate measures, which makes it difficult to evaluate the efficacy of such training in relation to its stated systems-level goals. Until researchers can show that measures of human service providers' cultural competence are predictive of objective client outcomes, inferences about how DT within health care settings measures up to its goals are speculative.

Limitations and Recommendations

Cultural competence training has been articulated as important for providing equitable services. However, very few studies have examined the impact of cultural competence training on actual systems-level outcomes, such as quality of care for historically marginalized clients or disparities in treatment, morbidity, and mortality. Instead, many studies from the field determine the efficacy of cultural competence training by relying on surrogate measures of individual-level provider knowledge, awareness, and self-efficacy. As such, the rationale supporting a cultural competence approach remains circular and rests strongly on the theoretical benefits of cultural competence, rather than on rigorous empirical evidence with respect to client outcomes (Saha et al. 2013). Therefore, future researchers should assess the extent to which human service providers' cultural competence serves as a reasonable surrogate measure for equitable care.

Additional concerns abound regarding the field's investment in cultural competence as a model for delivering effective care to clients from historically marginalized groups in the absence of a thorough clarification of the concept. Cultural competence, as a construct and curriculum, largely eludes easy definitions or operationalizations (Ridley et al. 2001). As such, cultural competence training remains underspecified and represents a wide range of heterogenous practices, with little attention to the active ingredients of the approach (Sue 2001).

Beyond these concerns, we also encourage sensitivity to potential pitfalls in the approach, that, if not simultaneously attended to, could undermine the enterprise altogether. We caution against using cultural differences in a reductionist way to predict patient behavior and guide clinician-patient interactions. Such an approach can neglect the heterogeneity among members of cultural groups, encourage the use of race as a proxy for culture, and promote stereotyping. The group categorization processes that are necessary for considering clients' cultural background in treatment also pave the way for cultural stereotypes to inform health care decisions. The use of stereotypes in human service provision can bias the way providers perceive clients, lead them to be inattentive to individuating information, increase the attention to (and weighting

Table 2 Summary of the literature and our recommendations for DT in human service settings

DT in human services	
Goals and approach	■ The goal is to familiarize human service providers with culturally based beliefs to improve quality of care for historically marginalized clients and eliminate extant disparities in treatment outcomes. ■ The trainees engage in role playing, participatory learning, cultural immersion, and/or community-based practicums. ■ The training emphasizes both knowledge and skills relevant to culturally competent services.
Research designs and assessment of outcomes	■ Few studies (13.27%) used experimental designs, some (20.41%) used quasi-experimental designs, and a majority (66.33%) used pre–post designs. ■ Few studies (23.47%) included delayed assessment of outcomes. ■ Most studies (86.73%) utilized trainees' individual-level cognitive and affective reactions to the training as the primary outcome of interest. ■ Few studies (14.29%) examined outcomes that did not rely on self-report.
State of the evidence	■ Most studies found that DT in human services was associated with increases in provider-reported cultural competence. ■ There is conflicting and limited evidence regarding whether provider cultural competence is related to better care and services for members of historically marginalized groups. ■ Evidence is inconclusive regarding whether diversity-related programming reduces disparities in treatment outcomes or improves the experiences of historically marginalized clients.
Limitations	■ There is a mismatch between the goals of the training and the outcomes used to evaluate its efficacy (i.e., overreliance on surrogate or proxy measures). ■ Cultural competence, as a practice, is underspecified, largely theoretical, and not evaluated with respect to client outcomes.
Recommendations	■ Evaluate client-centered and systems-level outcomes, such as client satisfaction, quality of care, adherence to treatment recommendations, disparities in outcomes, and differences in morbidity and mortality. ■ Establish a relationship between provider-reported cultural competence and outcomes for clients from historically marginalized groups. ■ Pair cultural competence training with training on how to guard against the influence of stereotypes (e.g., seeking out individuating information).

Abbreviation: DT, diversity training.

of) stereotype-confirming information, and lead to implicit and nonverbal forms of bias in client-clinician interactions (Burgess et al. 2007, Stone & Moskowitz 2011).

Cultural competence training should be coupled with education and strategies regarding how to guard against the undue influence of stereotypes to mitigate biased decisions related to client care (Burgess et al. 2007, Pankey et al. 2018, Stone & Moskowitz 2011). The psychological literature concludes that, absent personal information about an individual, people often rely on stereotypes to make group-based generalizations. To combat these stereotypes, research suggests that human service providers should actively seek individuating information about a client to prevent stereotypes from filling in the gaps (e.g., Ehrke et al. 2014, Fiske & Neuberg 1990) (see **Table 2** for a summary of the literature and our recommendations for DT in human service settings).

DIVERSITY TRAINING FOR STUDENTS IN EDUCATIONAL SETTINGS

Schools are becoming increasingly diverse, yet students from historically marginalized groups still underperform in academic pursuits compared to students from majority groups and relative to their potential (Natl. Assess. Educ. Prog. 2015). The achievement gap is apparent across a

wide variety of educational outcomes, including standardized test scores, high school graduation rates, admission rates in secondary education, and placement in gifted and talented programs (Am. Psychol. Assoc. 2012). Although many factors likely contribute to the achievement gap, adverse school climates, which undermine feelings of belonging, may be partly responsible.

Walton & Cohen (2007) found that improving feelings of belonging on campus improved the course grades of historically marginalized students. In a more recent study (Murrar et al. 2020), students who reported being treated more inclusively by their peers had an increased sense of belonging and earned better grades. This research provides encouraging evidence to suggest that improving campus climate can promote the performance and retention of historically marginalized students in higher education.

Many scholars stress the importance of culturally relevant pedagogy as a way to promote a positive school climate and improve the experiences of students from historically marginalized groups within education. Numerous universities across the United States require some form of diversity and inclusion curricula, typically in the form of an ethnic studies course (Greens 2000). Diversity as a pedagogical requirement has not been without its critics (e.g., Goldstein 2019). These critiques, however, underscore the importance of understanding what content should be included in diversity-related curricula, whether diversity education is efficacious, and who reaps the benefits of diversity and inclusion in course content.

Goals and Approach

By scaffolding students' multicultural awareness, diversity and inclusion curricula are designed to improve school climate for members of historically marginalized groups and to mitigate disparities in educational outcomes. Therefore, in our review of the literature we asked, Do historically marginalized students report fewer instances of school-based discrimination, and do they perceive a greater sense of belonging in schools that require diversity-related coursework? Do schools that offer diversity-related programming demonstrate more equitable educational outcomes relative to schools that do not provide such programming?

We narrowed our focus to include articles that implemented curriculum-based DT for a general student audience rather than for student teachers or medical students, as discussed in the previous section on DT for human service providers. The majority of the studies under consideration delivered DT content during a semester-long college course that involved lectures, assignments, and small-group discussions. The topics of the courses varied; whereas some studies involved women studies courses (e.g., Case & Stewart 2010, Stake & Hoffmann 2001), others examined the impact of psychology of prejudice courses (e.g., Hogan & Mallott 2005, Kernahan & Davis 2010), and still others focused on human sexuality courses (e.g., Mansoori-Rostam & Tate 2017, McDermott et al. 2018).

Other scholars examined the impact of a briefer educational activity, such as a particular curriculum unit (McDermott et al. 2018) or experiential learning activity (Hillman & Martin 2002). Some researchers specifically emphasized the importance of intergroup learning through discussions with classmates from diverse backgrounds (e.g., Nagada et al. 2004, Schmidt et al. 2019). Our sample also included studies investigating the impact of panel presentations (McDermott et al. 2018), role playing (Hillman & Martin 2002), study abroad programs (Clarke et al. 2009), and community events that celebrated diverse cultures (Klak & Martin 2003).

Research Designs and Assessment of Outcomes

Our search yielded 61 articles on DT in educational settings. Of these studies, 51 systematically and quantitatively evaluated diversity-related programming in a higher education setting. Most

of these studies (52.94%) used a quasi-experimental design by comparing students in a course with diversity and inclusion pedagogy to students in a control course. Other studies used pre–post designs (33.33%), for example, by collecting data on the first and last days of the semester (e.g., Fischer 2010). Few studies (13.73%) used experimental designs. Although most studies assessed outcomes collected immediately following the training, some studies collected delayed outcomes (29.41%).

The primary outcome of interest for most studies (94.12%) were students' individual-level, cognitive, and affective reactions. For example, researchers examined students' racial attitudes (e.g., Hogan & Mallott 2005, Rudman et al. 2001), homophobia (e.g., Hillman & Martin 2002, Hodson et al. 2009), and sexism (e.g., Pettijohn & Walzer 2008, Yoder et al. 2016). Other studies examined participants' acknowledgment of heterosexual (Case & Stewart 2010), White (Cole et al. 2011), and male (Case 2007) privilege. Our sample also contained studies that evaluated the extent to which curriculum-based DT fostered awareness of diversity-related challenges, such as generalized cultural awareness (Fischer 2010), perceived gender equality (Colvin-Burque et al. 2007), and awareness of racism (Cole et al. 2011).

Very few studies (3.92%) examined outcomes beyond the trainees' self-reported attitudes and knowledge. In rare exceptions, the researchers examined observational data from classroom discussions (Ross 2014) and changes in students' measures of implicit bias (Rudman et al. 2001).

Many studies evaluated individual differences or contextual factors that may moderate the impact of diversity-related course material on students' attitudes. Some studies focused on student characteristics such as open-mindedness (Fischer 2010), race and empathy (Cole et al. 2011), need for cognition (Hogan & Mallott 2005), and course engagement (Pettijohn & Walzer 2008). In contrast, Rudman and colleagues (2001) evaluated the impact of interacting with a Black professor on students' attitudes both with and without the provision of diversity-related instruction.

State of the Evidence

Many studies demonstrated a reduction in students' self-reported prejudice following diversity-related coursework, relative to pretest scores at the beginning of the semester (e.g., Chang 2002, Colvin-Burque et al. 2007). Similarly, of the studies that utilized quasi-experimental designs, many found that students who completed a diversity education course reported lower levels of prejudice immediately following the course (e.g., Hussey et al. 2010, Rudman et al. 2001) compared to students enrolled in courses without diversity-related content.

When implementing a quasi-experimental design, however, researchers should be mindful of the possibility of sample bias among students who enroll in diversity-related courses. Although many students take diversity-related coursework as part of their ethnic studies requirement, other students may take these courses because of their interest in cultural differences, experiences with bias, or passion for diversity-related initiatives (Denson 2009). As a result, students who opt into diversity-related courses may possess traits that foster more positive change relative to students in comparison courses (Case 2007, Mansoori-Rostam & Tate 2017). In contrast, those required to take such courses may show backlash effects, which undercut the goal of promoting inclusion and reducing biases (Brannon et al. 2018, Vianden 2018). Researchers should be attentive to, and control for, potential self-selection biases in participant samples who enroll in diversity-related courses, and they should guard against the potential adverse effects of requiring participation in these courses.

In addition, pre–post assessments may be particularly vulnerable to demand effects. Demand effects refer to biased findings that occur when participants infer the experimenters' hypothesis and then respond to measures in a way that tends to confirm the researchers' prediction (Weber &

Cook 1972). Research suggests that demand effects can be exacerbated if participants have positive attitudes toward the experimenter (Nichols & Maner 2008), which is likely true of many studies in which the experimenter administering questionnaires is also the course's instructor.

As was true in organizational and human service settings, relatively few studies within education quantitatively evaluated the long-term impact of diversity-related programming. When delayed assessments were included, the evidence was decidedly mixed. For example, Hogan & Mallott (2005) observed reduced homophobia immediately following a psychology of prejudice course, but the effect did not persist across semesters. More encouragingly, McDermott and colleagues (2018) evaluated the impact of a panel presentation and trans-themed film and found a reduction in self-reported prejudice that persisted when evaluated 6 weeks later.

Ethnic studies courses and diversity-related education have been proposed as a panacea for greater inclusion on college campuses. However, caution is warranted given the mixed evidence of the long-term efficacy of ethnic studies courses on cognitive measures. In addition, researchers' use of self-reported attitudes and knowledge is silent on the extent to which DT in school settings measures up to its stated systems-level goals.

Limitations and Recommendations

Consistent with our findings in other contexts, our review of the literature on diversity-related curricula in educational settings found that researchers relied on students' individual-level attitudes as surrogate outcomes for measuring inclusive campus climate and concluded that trainings were effective without considering the perspectives of students from historically marginalized groups. We recommend that future work examine changes over time in systems-level outcomes—such as perceptions of school climate, disparities in academic achievement, and historically marginalized students' sense of belonging—as more appropriate tests of the benefits of DT in educational settings.

Furthermore, diversity courses that target students' knowledge and awareness without attending to mechanisms of behavioral change are likely not sufficient to create lasting changes in the form of reduced expressions of bias, increased intergroup inclusion, and improved feelings of belonging for marginalized students. Of the studies reviewed, only one study (Pedersen & Barlow 2008) explicitly implemented antiprejudice strategies throughout the course of an educational program; these researchers used tactics such as combating false beliefs, invoking empathy, meeting local needs, and focusing on changing behaviors as much as attitudes.

Incorporating evidence-based prejudice reduction strategies alongside diversity-related course content is likely essential if DT seeks to make meaningful changes in discriminatory behaviors, not just attitudes (Sanchez & Medkik 2004). Given that one of the goals of DT on college campuses is to promote greater inclusion among the student body, educators should strive to do more than teaching students diversity-related content. To achieve the goal of inclusion for members of historically marginalized groups, curriculum-based DT should also take advantage of bias reduction and inclusion-promoting strategies identified as effective in the psychological literature.

Consider, for example, how educational settings can more intentionally target prejudice-related behaviors through social norm change. Although targeted social norm communication can be leveraged effectively in a variety of settings, pressures to belong and conform are amplified in young adults, rendering social norms particularly salient on college campuses. With this in mind, we echo the recommendations of others in suggesting that targeted social norm communication can be a powerful approach for reducing prejudicial behavior within school settings (e.g., Murrar et al. 2020, Tankard & Paluck 2016). Higher education institutions can implement social norm change through the communication of diversity-related values in the form of pro-diversity posters (Murrar et al. 2020), campus events (Klak & Martin 2003), and the recruitment of students from

marginalized groups (Hurtado 2005). Fellow students can be particularly influential in the communication of social norms through the confrontation of prejudice (Czopp & Monteith 2003), student-led protests and organizations (Paluck et al. 2016), and diversity-related discussions both in and out of the classroom (Alimo 2012).

Another prejudice reduction strategy from the social psychological literature that may be particularly fruitful within the context of DT for higher education is purposeful intergroup contact. Classroom settings naturally create ideal contexts (e.g., small group discussions, collaborative group projects) in which intergroup contact can meet the requirements needed for reductions in bias (Allport 1954). In a meta-analysis of 16 studies, Denson (2009) found that ethnic studies courses that provided students with additional positive interracial contact produced larger effects on students' attitudes compared to courses that just focused on teaching diversity-related course content. This research suggests that direct intergroup contact with peers from diverse backgrounds may be important for providing students with a space to apply the content gleaned from their courses, as this enables them to actualize their more positive intergroup attitudes into more inclusive behaviors (Gurin et al. 2004, Zúñiga et al. 2002).

However, in considering the utility of intergroup contact as a tool for increased inclusion on college campuses, further attention should be allocated to the experiences of contact for people of color. Some research suggests that the positive effects of intergroup contact may not extend to members of historically marginalized groups (Dixon et al. 2010, Pettigrew & Tropp 2006, Schellhaas & Dovidio 2016). Given that the ultimate goal of DT in education is to improve the experiences of individuals from historically marginalized groups, future researchers should prioritize evaluating the extent to which intergroup contact improves not only the attitudes of majority group members but also the experiences of marginalized group members. To this end, forthcoming research should ensure that individuals from historically marginalized groups are not overburdened facilitators of intergroup contact without benefit (see **Table 3** for a summary of the literature and our recommendations for DT in educational settings).

RECOMMENDATIONS FOR DIVERSITY TRAINING ACROSS CONTEXTS

We began our deep dive into the literature on DT with the intention of illuminating best practices in the delivery of DT. We discovered that the available scholarship on DT is large and complex, and there is little consensus on the specific definitions of DT, the overarching goals of DT, or the particular practices that comprise DT. Further complicating these efforts, our review of the empirical literature in each discipline revealed a number of shortcomings that limit our ability to draw clear conclusions regarding which, if any, DT programs are effective in reaching their objectives. More troubling, many studies reveal the potential for adverse effects following DT (e.g., Brannon et al. 2018, Legault et al. 2011). Unfortunately, our primary conclusion following our review of the recent literature echoes that of scholars who conducted reviews of the DT literature in the past. Despite multidisciplinary endorsement of the practice of DT, we are far from being able to derive clear and decisive conclusions about what fosters inclusivity and promotes diversity within organizations (Bezrukova et al. 2016, Carter et al. 2020, Green & Hagiwara 2020, Moss-Racusin et al. 2014, Paluck 2006, Paluck & Green 2009, Paluck et al. 2020).

This state of affairs is concerning, particularly in light of the enthusiasm for, and monetary investment in, DT. Implementation of DT has clearly outpaced the available evidence that such programs are effective in achieving their goals. As such, we advise caution and tempered enthusiasm for the widespread implementation of DT. In the spirit of propelling the science of DT forward, we offer some recommendations, which cut across contexts or subfields, for how to

Table 3 Summary of the literature and our recommendations for DT in educational settings

DT in educational settings	
Goals and approach	■ The goal is to increase belonging, retention, and achievement of students from historically marginalized groups and promote an inclusive school climate. ■ The training most often consists of semester-long courses (e.g., on psychology of prejudice, women studies, human sexuality, etc.). ■ Students are educated on the complex histories and perspectives of members from historically marginalized groups.
Research designs and assessment of outcomes	■ Few studies (13.73%) used experimental designs, most (52.94%) used quasi-experimental designs, and some (33.33%) used pre–post designs. ■ Few studies (29.41%) evaluated the long-term impacts of diversity-related programming. ■ Most studies (94.12%) evaluated students' individual-level attitudes, beliefs, and knowledge as the primary outcomes of interest. ■ Very few studies (3.92%) examined outcomes that did not rely on self-report.
State of the evidence	■ Most studies found that DT in college settings was associated with more positive attitudes toward members of historically marginalized groups at the end of the semester. ■ There is conflicting and limited evidence regarding the long-term effects of DT for students in educational settings. ■ Evidence is inconclusive regarding whether diversity-related programming in education improves school climate and achievement for students from historically marginalized groups.
Limitations	■ There is a mismatch between the goals of the training and the outcomes used to evaluate its efficacy (i.e., overreliance of surrogate or proxy measures). ■ Concerns regarding self-selection bias, backlash effects, and demand effects arise when interpreting the results of prior studies.
Recommendations	■ Evaluate systems-level outcomes, such as changes in perceptions of school climate, disparities in academic achievement, and historically marginalized students' sense of belonging. ■ Attend to self-selection bias in quasi-experimental designs, backlash effects in required DT, and demand effects in pre–post designs. ■ Incorporate evidence-based prejudice reduction strategies (e.g., social norm communication, intergroup contact, etc.) alongside diversity-related course content.

Abbreviation: DT, diversity training.

build a more rigorous and relevant science of DT. Following these recommendations would allow DT scientists to create an evidence base that would have clear and applied utility for DT practitioners and consumers. We encapsulated the implications of these recommendations into a list of questions that every scholar of DT should be able to answer and every practitioner and consumer of DT should want to know before implementing any given DT (see **Table 4**). Although the list is not exhaustive, it is our hope that readers will use **Table 4** to ask hard questions about DT programs and to be more purposeful in the selection, implementation, and evaluation of DT across contexts.

First, no single DT should be marketed as a magic bullet for equity. Rather, we advocate for an approach to the development of DT that is grounded in relevant theory and informed by empirical evidence to justify the content of the training, the rationale for the practices used, the boundary conditions anticipated, and the hypothesized mechanisms by which the program effects change. It is not enough to simply consult prior research and pluck bias reduction strategies from the literature, such as individuation, social norm communication, or intergroup contact. The elements of DT should be selected to align with the particular goals of the organization and to address the specific problem the program is designed to solve. Although existing psychological research can provide practitioners of DT with hypotheses about how bias reduction strategies may operate

Table 4 Recommended questions for proponents of DT to consider[a]

Guiding questions for DT researchers, practitioners, and consumers	
What are the goals in implementing this training?	■ What is the problem to be solved with DT? ■ How will we know if the DT program is effective? ■ What are the ultimate and surrogate (i.e., proxy) outcomes of interest?
What content is being presented within the intervention?	■ Does the content of the training reflect the goals to be achieved? ■ What is the theoretical and empirical basis for the training? ■ Does the DT program include concrete action recommendations in the service of accomplishing the goals of the training?
What is the evidence to support this DT program?	■ Are the samples used in prior research relevant to the goals of implementation? ■ How does the evidence in support of the DT program's effects bear on the systems-level goals guiding implementation? ■ Have the outcomes evaluated been consequential? Have the effects been long-lasting? ■ Are there any potentially adverse effects of the DT that should be guarded against? ■ Does the DT program include a plan for the ongoing investigation of the program's efficacy?
Is the context of implementation promoting or impeding the intervention's efficacy?	■ Are experts being utilized to develop, deliver, and evaluate the DT program? ■ Is the DT program being implemented alongside other organizational diversity initiatives and policies? ■ Does the DT program attend to the perspectives and experiences of individuals from historically marginalized groups?

Abbreviation: DT, diversity training.

[a]We encourage scientists to use this table to guide the development, implementation, and evaluation of DT studies to build a more rigorous and relevant diversity science. Practitioners should be able to respond to these inquiries when delivering and endorsing any program of DT. Consumers, armed with these questions, can increase their confidence in the efficacy of a particular DT for improving inclusion and equity within their organization.

within a specific DT, these hypotheses need to be assessed within the context of the full training and in reference to the long-term goals of the program. In proposing a systematic approach to the creation of DT programs, we recommend that programs be tailored to the specific context of implementation and revised in an iterative fashion, based on evidence, to enhance their efficacy.

Next, we suggest that DT research needs to become more rigorous. As noted previously, some scholars argue that we should be ethically bound to demonstrate that DT programs are effective and, just as important, do no harm (Paluck 2012). To this end, we advocate for the use of experimental designs with relevant samples to provide persuasive evidence of the utility of a particular approach. Whenever possible, it is recommended that researchers undertake the challenge of randomized controlled trials to provide causal evidence of the hypothesized effects of a DT. In the best possible circumstances, control groups will contain an active component that will enable researchers to test the efficacy of a particular DT program against an alternative training program. When not viable, wait-list controls can be utilized to test the short-term effects of DT and to ensure that a particular DT, if found to be effective, is eventually disseminated to all members of an organization. Whenever randomized controlled trials are not feasible, researchers should stay vigilant to, and control for, potential threats to the validity of their studies. In addition, experimental studies should extend outside the lab, and into the field settings with relevant populations, to evaluate the functionality of a particular approach for practitioners and consumers (see also Paluck et al. 2020).

In developing a more rigorous and relevant science of DT, greater attention should be paid to the types of outcomes that will provide evidence that the DT offered is actually effective. This process starts with an analysis of the particular problem an organization is trying to solve by implementing DT. This analysis should then determine the goals for training, the relevant DT

approach, and the outcomes that will reveal if the DT was effective (Campbell & Brauer 2020, Carter et al. 2020). Failing to undertake this type of analysis limits the utility of DT research.

One of the most striking features of current research on DT across disciplines is that the outcomes most often used to examine the impact of DT are, at best, limited in terms of what they can reveal regarding the efficacy of the training. We observed throughout our review an overreliance on surrogate and individual-level measures, all of which could be helpful in achieving goals related to equity, yet few of them bear directly on the stated goals of the DT. To better advance the science of DT, scholars must hold the success of their interventions to a higher standard by attending to more than just individual-level self-reported outcomes. We encourage future researchers to extend beyond commonly used cognitive and affective measures and to assess instead a wide range of outcomes (including consequential, behavioral, and systems-level outcomes) to better shed light on the potential breadth of the effects of a DT program.

DT is marketed to improve the experiences of employees, clients, and students from historically marginalized groups and to achieve greater equity across settings. However, research on DT, as well as its practice, only infrequently attends to the perspectives and experiences of individuals who are at risk for experiencing discrimination. Given this focus, we suggest that historically marginalized individuals should be consulted during the planning process, if DT is to be effective in meeting its goals. Input and involvement from members of historically marginalized groups should be actively sought in determining whether and how to deliver DT within their settings. By ensuring that historically marginalized individuals have a seat at the table in the development and selection of diversity-related initiatives, organizations can certify that these individuals' voices are heard and their perspectives are represented. Beyond involving members of historically marginalized groups in the early planning stages of DT, evidence supporting the efficacy of DT necessitates a better understanding of the experiences of historically marginalized individuals as a function of diversity, equity, and inclusion programming (Roberts et al. 2020).

Finally, researchers should better determine the long-term impacts of their trainings by conducting follow-up assessments over time. Rather than implementing and evaluating DT within a "one-and-done" approach, the science of DT would be better served by longitudinal assessments of systems-level outcomes to ensure that DT achieves its stated goals. Specifically, we argued for an evaluation of the recruitment, promotion, and retention of employees from historically marginalized groups in organizational settings. We asked researchers in human services to attend to rates of morbidity, mortality, and treatment adherence in validating the benefits of cultural competence training. And we advised scholars of diversity education not to neglect perceived belonging, retention, and achievement of students from historically marginalized groups in higher education.

In advancing these recommendations, we acknowledge that these recommendations impose significant demands on the proponents of DT. The kinds of studies we are advocating represent enormous undertakings that would require tremendous resources in terms of time, money, and personnel. Considering these challenges, it may be easy to understand why many studies to date have involved less rigorous empirical methods and easy-to-collect outcome measures. However, given the immense investment in the practice of DT, a more ambitious research agenda is necessary for DT to measure up to its stated goals. In the service of ensuring that diversity scholarship offers utility to both practitioners and consumers, particularly for individuals facing discrimination, we cannot be complacent about the current state of the evidence in support of DT.

In the pursuit of an improved science of DT, we join other scholars (Carter et al. 2020, Paluck 2006) who suggest that building this type of rigorous and relevant evidentiary base necessitates the forging of collaborative relationships between social scientists and organizations. Partnerships between scholars and organizations will allow research on DT to strike a balance between scientific rigor and practical utility by providing future researchers with a potential avenue for bringing

their work out of the lab and into settings where the benefits of psychological research can be realized. Through these collaborations, practitioners can be better grounded in the empirical literature, and DT research can become better contextualized in applied settings. Supporting this suggestion, we note that some of the most methodologically impressive studies in this review are the result of successful partnerships (e.g., Carnes et al. 2015, Chang et al. 2019, Moss-Racusin et al. 2018). Although field experiments of this kind require careful consideration on the part of both practitioners and researchers, such partnerships would be mutually beneficial and, importantly, would offer the best way to ensure that DT lives up to consumer expectations and rigorous experimental standards.

In closing, we acknowledge that our review could be viewed as a general rebuke of DT as an enterprise. We would like to clarify that this is not at all the case. Indeed, we are strong proponents of the importance of creating conditions whereby members of historically marginalized groups can be included, feel respected, and thrive. We appreciate the potential benefits of diversity-related initiatives and value the goals upon which the practice of DT has been built. We urge the field to create a more rigorous and relevant science of DT in the service of making it possible for the practice of DT to achieve its goals. We recognize that the challenge we laid out for the science of DT is enormous and echoes, in many ways, the calls advanced in prior reviews of DT (e.g., Moss-Racusin et al. 2014, Paluck & Green 2009, Paluck et al. 2020). The enormity of the challenge, however, pales in comparison to the potential benefits of DT. We hope that our call to action and the roadmap for how to build a better diversity science will make it possible for the next review of the DT literature to offer effective, evidence-based best practices in DT.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank undergraduate members of the University of Wisconsin–Madison's Prejudice and Intergroup Relations Lab for their work collecting and coding articles for this review. In addition, we appreciate the thoughtful feedback received from Megan Bruun, Emily Dix, Katharine Scott, and Katherine Swerbenski on earlier versions of this manuscript. In particular, we would like to thank Nicole Huth for her substantial contributions to the project and her helpful comments on earlier drafts of the manuscript. Preparation of this review was also supported by the National Institutes of Health (grant R011R35GM127043–01). The opinions expressed are those of the authors and do not represent the views of the National Institutes of Health.

LITERATURE CITED

Abernethy AD. 2005. Increasing the cultural proficiency of clinical managers. *J. Multicult. Couns. Dev.* 33:81–93

Alimo CJ. 2012. From dialogue to action: the impact of cross-race intergroup dialogue on the development of White college students as racial allies. *Equity Excell. Educ.* 45(1):36–59

Allport G. 1954. *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley

Amatea ES, Chloewa B, Mixon KA. 2012. Influencing preservice teachers' attitudes about working with low-income and/or ethnic minority families. *Urban Educ.* 47(4):801–34

Am. Psychol. Assoc. 2012. *Ethnic and racial disparities in education: psychology's contributions to understanding and reducing racial disparities*. Rep., Pres. Task Force Educ. Disparities, Am. Psychol. Assoc., Washington, DC. **https://www.apa.org/ed/resources/racial-disparities.pdf**

Amerson R. 2010. The impact of service-learning on cultural competence. *Nurs. Educ. Res.* 31(1):18–22

Beach MC, Price EG, Gary TL, Robinson KA, Gozu A, et al. 2005. Cultural competence: a systematic review of health care provider educational interventions. *Med. Care* 43(4):356–73

Beagan BL. 2003. Teaching social and cultural awareness to medical students: "It's all very nice to talk about it in theory, but ultimately it makes no difference." *Acad. Med.* 76(6):605–14

Berlin A, Nilsson G, Törnkvist L. 2010. Cultural competence among Swedish child health nurses after specific training: a randomized trial. *Nurs. Health Sci.* 12:381–91

Bezrukova K, Spell CS, Perry JL, Jehn KA. 2016. A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychol. Bull.* 142(11):1227–74

Bhattacharjee R. 2016. 11 women kicked off Napa Valley Wine Train for laughing too loudly settle racial discrimination suit. *NBC Bay Area*, April 18

Brannon TN, Carter ER, Murdock-Perriera L, Higginbotham GD. 2018. From backlash to inclusion for all: instituting diversity efforts to maximize benefits across group lines. *Soc. Iss. Policy Rev.* 12(1):57–90

Brathwaite AC, Majumdar B. 2006. Evaluation of a cultural competence educational programme. *J. Adv. Nurs.* 53(4):470–79

Brewer MB, Hippel WV, Gooden MP. 1999. Diversity and organizational identity: the problem of entrée after entry. In *Cultural Divides: Understanding and Overcoming Group Conflict*, ed. DA Prentice, DT Miller, pp. 337–63. New York: Russell Sage Found.

Brochu PM. 2020. Testing the effectiveness of a weight bias educational intervention among clinical psychology trainees. *J. Appl. Soc. Psychol.* In press. **https://doi.org/10.1111/jasp.12653**

Burgess D, van Ryn M, Dovidio J, Saha S. 2007. Reducing racial bias among health care providers: lessons from social-cognitive psychology. *J. Gen. Intern. Med.* 22(6):882–87

Bush VD, Ingram TN. 2001. Building and assessing cultural diversity skills: implications for sales training. *Ind. Mark. Manag.* 30(1):65–76

Campbell MR, Brauer M. 2020. Incorporating social-marketing insights into prejudice research: advancing theory and demonstrating real-world applications. *Perspect. Psychol. Sci.* 15(3):608–29

Carnes M, Devine PG, Manwell LB, Byars-Winston A, Fine E, et al. 2015. Effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. *Acad. Med.* 90(2):221–30

Carratala S, Maxwell C. 2020. *Health disparities by race and ethnicity*. Fact Sheet, Cent. Am. Prog., Washington, DC. **https://cdn.americanprogress.org/content/uploads/2020/05/06130714/HealthRace-factsheet.pdf**

Carter ER, Onyeador IN, Lewis NA. 2020. Developing and delivering effective anti-bias training: challenges and recommendations. *Behav. Sci. Policy* 6(1):57–73

Carter MM, Lewis EL, Sbrocco T, Tanenbaum R, Oswald JC, et al. 2006. Cultural competency training for third-year clerkship students: effects of an interactive workshop on student attitudes. *J. Natl. Med. Assoc.* 98(11):1772–78

Case KA. 2007. Raising male privilege awareness and reducing sexism: an evaluation of diversity courses. *Psychol. Women Q.* 31(4):426–35

Case KA, Stewart B. 2010. Changes in diversity course student prejudice and attitudes toward heterosexual privilege and gay marriage. *Teach. Psychol.* 37(3):172–77

Castillo LG, Brossart D, Conoley CW. 2007. The influence of multicultural training on perceived multicultural counseling competencies and implicit racial prejudice. *J. Multcult. Couns. Dev.* 35:243–55

Castro A, Ruiz E. 2009. The effects of nurse practitioner cultural competence on Latina patient satisfaction. *J. Am. Acad. Nurs. Pract.* 21(5):278–86

Chang EH, Milkman KL, Gromet DM, Rebele RW, Massey C, et al. 2019. The mixed effects of online diversity training. *PNAS* 116(16):7778–83

Chang MJ. 2002. The impact of an undergraduate diversity course requirement on students' racial views and attitudes. *J. Gen. Educ.* 15(1):21–42

Clarke I, Flaherty TB, Wright ND, McMillen RM. 2009. Student intercultural proficiency from study abroad programs. *J. Mark. Educ.* 31(2):173–81

Cole ER, Case KA, Rios D, Curtin N. 2011. Understanding what students bring to the classroom: moderators of the effects of diversity courses on student attitudes. *Cult. Divers. Ethn. Minor. Psychol.* 17(4):397–405

Colvin-Burque A, Zugazaga CB, Davis-Maye D. 2007. Can cultural competence be taught? Evaluating the impact of the soap model. *J. Soc. Work Educ.* 43(2):223–42
Combs GM, Luthans F. 2007. Diversity training: analysis of the impact of self-efficacy. *Hum. Res. Dev. Q.* 18(1):91–120
Cox WTL, Devine PG. 2019. The prejudice habit-breaking intervention: an empowerment-based confrontation approach. In *Confronting Prejudice and Discrimination*, ed. RK Mallett, Monteith, pp. 249–74. San Diego, CA: Elsevier
Crandall SJ, George G, Marion GS, Davis S. 2003. Applying theory to the design of cultural competency training for medical students: a case study. *Acad. Med.* 78(6):588–94
Crespo G. 2018. Doctor says she was racially profiled while trying to help fellow passenger. *CNN*, Nov. 1
Czopp AM, Monteith MJ. 2003. Confronting prejudice (literally): reactions to confrontations of racial and gender bias. *Pers. Soc. Psychol. Bull.* 29(4):532–44
Denson N. 2009. Do curricular and cocurricular diversity activities influence racial bias? A meta-analysis. *Rev. Educ. Res.* 79(2):805–38
Devine PG. 1989. Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56(1):5–18
Devine PG, Forscher PS, Cox WTL, Kaatz A, Sheridan J, Carnes M. 2017. A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *J. Exp. Soc. Psychol.* 73:211–15
Diaz-Lazaro CM, Cohen BB. 2001. Cross-cultural contact in counseling training. *J. Multicult. Couns. Dev.* 29(1):41–56
Dixon J, Durrheim K, Tredoux C, Tropp LR, Clack B, Eaton E. 2010. A paradox of integration? Interracial contact, prejudice reduction and black South Africans' perceptions of racial discrimination. *J. Soc. Issues* 66:403–18
Dobbin F, Kalev A. 2016. Why diversity programs fail. *Harv. Bus. Rev.* 94(7):52–60
Dogra N. 2001. The development and evaluation of a programme to teach cultural diversity to medical undergraduate students: teaching cultural diversity. *Med. Educ.* 35(3):232–41
Dover TL, Kaiser CR, Major B. 2020. Mixed signals: the unintended effects of diversity initiatives. *Soc. Issues Policy Rev.* 14(1):152–81
Ehrke F, Berthold A, Steffens MC. 2014. How diversity training can change attitudes: increasing perceived complexity of superordinate groups to improve intergroup relations. *J. Exp. Soc. Psychol.* 53:193–206
Fischer R. 2010. Cross-cultural training effects on cultural essentialism beliefs and cultural intelligence. *Int. J. Intercult. Relat.* 35(6):767–75
Fiske ST, Neuberg SL. 1990. A continuum of impression formation, from category-based to individuating processes: influence of information and motivation on attention and interpretation. In *Advances in Social Psychology*, ed. MP Zanna, pp. 1–74. Cambridge, MA: Academic
Flores G. 2000. Culture and the patient-physician relationship: achieving cultural competency in health care. *J. Pediatr.* 136(1):14–23
Forscher PS. 2017. *The individually-targeted habit-breaking intervention and group-level change*. PhD Thesis, Univ. Wisconsin–Madison
Forscher PS, Lai CK, Axt JR, Ebersole CR, Herman M, et al. 2019. A meta-analysis of procedures to change implicit measures. *J. Pers. Soc. Psychol.* 117(3):522–59
Goldstein D. 2019. Push for ethnic studies in schools faces a dilemma: whose stories to tell. *New York Times*, Aug. 15
Green TL, Hagiwara N. 2020. The problem with implicit bias training. *Scientific American*, Aug. 28
Greens E. 2000. Most colleges require diversity education. *Chron. High. Educ.* 47(10):A16
Griswold K, Kernan JB, Servoss TJ, Saad FG, Wagner CM, Zayas LE. 2006. Refugees and medical student training: results of a programme in primary care. *Med. Educ.* 40(7):697–703
Gurin P, Nagada BA, Lopez GE. 2004. The benefits of diversity in education for democratic citizenship. *J. Soc. Issues* 60(1):17–34
Henderson S, Kendall E, See L. 2011. The effectiveness of culturally appropriate interventions to manage or prevent chronic disease in culturally and linguistically diverse communities: a systematic literature review. *Health Soc. Care Community* 19(3):225–49

Hennes EP, Pietri ES, Moss-Racusin CA, Mason KA, Dovidio JF. 2018. Increasing the perceived malleability of gender bias using a modified video intervention for diversity in STEM (VIDS). *Group Proc. Intergroup Relat.* 21(5):788–809

Hill ME, Augoustinos M. 2001. Stereotype change and prejudice reduction: short- and long-term evaluation of a cross-cultural awareness programme. *J. Community Appl. Soc. Psychol.* 11:243–62

Hillman J, Martin RA. 2002. Lessons about gay and lesbian lives: a spaceship exercise. *Teach. Psychol.* 29(4):308–11

Hodson G, Choma BL, Costello K. 2009. Experience alien-nation: effects of a simulation intervention on attitudes toward homosexuals. *J. Exp. Soc. Psychol.* 45(4):974–78

Hogan DE, Mallott M. 2005. Changing racial prejudice through diversity education. *J. Coll. Stud. Dev.* 46(2):115–25

Holladay CL, Quiñones MA. 2008. The influence of training focus and trainer characteristics on diversity training effectiveness. *Acad. Manag. Learn. Educ.* 7(3):343–54

Holman EG, Laundy-Meyer L, Fish JN. 2020. Creating supportive environments for LGBT older adults: an efficacy evaluation of staff training in a senior living facility. *J. Gerontol. Soc. Work* 65(5):464–77

Homan AC, Buengeler C, Eckhoff RA, van Ginkel WP, Voelpel SC. 2015. The interplay of diversity training and diversity beliefs on team creativity in nationally diverse teams. *J. Appl. Psychol.* 100(5):1456–67

Hood JN, Muller HJ, Seitz P. 2001. Attitudes of Hispanics and Anglos surrounding a workforce diversity intervention. *Hisp. J. Behav. Sci.* 23(4):444–58

Hostager TJ, De Meuse KP. 2008. The effects of a diversity learning experience on positive and negative diversity perceptions. *J. Bus. Psychol.* 23(3–4):127–39

Hughes KH, Hood LJ. 2007. Teaching methods and an outcome tool for measuring cultural sensitivity in undergraduate nursing students. *J. Transcult. Nurs.* 18(1):57–62

Hurtado S. 2005. The next generation of diversity and intergroup relations research. *J. Soc. Issues* 61(3):595–610

Hussey HD, Felck BKB, Warner RM. 2010. Reducing student prejudice in diversity-infused core psychology classes. *Coll. Teach.* 58(3):85–92

Jackson SM, Hillard AL, Schneider TR. 2014. Using implicit bias training to improve attitudes toward women in STEM. *Soc. Psychol. Educ.* 17(1):419–38

Johnson LM, Antle BF, Barbee AP. 2009. Addressing disproportionality and disparity in child welfare: evaluation of an anti-racism training for community service providers. *Child. Youth Serv. Rev.* 31:688–96

Juarez JA, Marvel K, Brezinski KL, Glazner C, Towbin MM, Lawton S. 2006. Bridging the gap: a curriculum to teach residents cultural humility. *Fam. Med.* 38(2):97–102

Kagnici DY. 2014. Reflections of a multicultural counseling course: a qualitative study with counseling students and counselors. *Educ. Sci. Theory Pract.* 14(1):53–62

Kalinonski ZT, Steele-Johnson D, Peyton EJ, Leas KA, Steinke J, Bowling NA. 2013. A meta-analytic evaluation of diversity training outcomes. *J. Organ. Behav.* 34(8):1076–104

Kernahan C, Davis T. 2010. What are the long-term effects of learning about racism? *Teach. Psychol.* 37(1):41–45

Klak T, Martin P. 2003. Do university-sponsored international cultural events help students to appreciate "difference"? *Int. J. Intercult. Relat.* 27(4):445–65

Kulik CT, Pepper MB, Roberson L, Parker SK. 2007. The rich get richer: predicting participation in voluntary diversity training. *J. Organ. Behav.* 28(6):753–69

LaPiere RT. 1934. Attitudes versus actions. *Soc. Force* 13:230–37

Legault L, Gutsell JN, Inzlicht M. 2011. Ironic effects of antiprejudice messages: how motivational interventions can reduce (but also increase) prejudice. *Psychol. Sci.* 22(12):1472–77

Leyva VL, Breshears EM, Ringstad R. 2014. Assessing the efficacy of LGBT cultural competency training for aging services providers in California's central valley. *J. Gerontol. Soc. Work* 57:335–48

Lie DA, Lee-Rey E, Gomez A, Bereknyei S, Braddock CH. 2011. Does cultural competency training of health professionals improve patient outcomes? A systematic review and proposed algorithm for future research. *J. Gen. Intern. Med.* 26(3):317–25

Lipman J. 2018. How diversity training infuriates men and fails women. *Time*, Jan. 25

LoboPrabhu S, King C, Albucher R, Liberzon I. 2000. A cultural sensitivity training workshop for psychiatry residents. *Acad. Psychiatry* 24(2):77–84
Luger L. 2011. Enhancing cultural competence in staff working with people with drug and alcohol problems—a multidimensional approach to evaluating the impact of education. *Soc. Work Educ.* 30(2):223–35
Madera JM, Neal JA, Dawson M. 2011. A strategy for diversity training: focusing on empathy in the workplace. *J. Hosp. Tour. Res.* 35(4):469–87
Majumdar B, Browne G, Roberts J, Carpio B. 2004. Effects of cultural sensitivity training on health care provider attitudes and patient outcomes. *J. Nurs. Scholarsh.* 36(2):161–66
Mansoori-Rostam SM, Tate CC. 2017. Peering into the "black box" of education interventions and attitude change: Audience characteristics moderate the effectiveness. . .and then only toward specific targets. *J. Soc. Psychol.* 157(1):1–15
Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. 2002. A demonstration of the impact of response bias on the results of patient satisfaction surveys. *Health Serv. Res.* 37(5):1403–17
McDermott DT, Brooks AS, Rohleder P, Blair K, Hoskin RA, McDonagh LK. 2018. Ameliorating transnegativity: assessing the immediate and extended efficacy of a pedagogic prejudice reduction intervention. *Psychol. Sex.* 9(1):69–85
Middleton VA. 2002. Increasing preservice teachers' diversity beliefs and commitment. *Urban Rev.* 34(4):343–61
Moss-Racusin CA, Pietri ES, Hennes EP, Dovidio JF, Brecoll VL, et al. 2018. Reducing STEM in gender bias with VIDS (video interventions for diversity in STEM). *J. Exp. Soc. Psychol. Appl.* 24(2):236–60
Moss-Racusin CA, van der Toorn J, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. 2014. Scientific diversity interventions. *Science* 343(6171):615–16
Murrar S, Campbell M, Brauer M. 2020. Exposures to peers' pro-diversity attitudes increases inclusion and reduces the achievement gap. *Nat. Hum. Behav.* 4:889–97
Nagada BA, Kim C, Truelove Y. 2004. Learning about difference, learning with others, learning to transgress. *J. Soc. Issues* 60(1):195–214
Natl. Assess. Educ. Prog. 2015. *School composition and the Black–White achievement gap*. Rep., Natl. Cent. Educ. Stat., US Dep. Educ., Washington, DC
Nichols AL, Maner JK. 2008. The good-subject effect: investigating participant demand characteristics. *J. Gen. Psychol.* 135(2):151–66
Paluck EL. 2006. Diversity training and intergroup contact: a call to action research. *J. Soc. Issues* 62(3):577–95
Paluck EL. 2012. Interventions aimed at the reduction of prejudice and conflict. In *The Oxford Handbook of Intergroup Conflict*, ed. LR Tropp, pp. 179–92. New York: Oxford Univ. Press
Paluck EL, Green DP. 2009. Prejudice reduction: What works? A review and assessment of research and practice. *Annu. Rev. Psychol.* 60:339–67
Paluck EL, Porat R, Clark CS, Green DP. 2020. Prejudice reduction: progress and challenges. *Annu. Rev. Psychol.* 72:533–60
Paluck EL, Shepherd H, Aronow PM. 2016. Changing climates of conflict: a social network experiment in 56 schools. *PNAS* 113(3):566–71
Pankey T, Alexander L, Carnes M, Kaatz A, Kolemainen C, et al. 2018. Breaking the bias-habit: a workshop to help internal medicine residents reduce the impact of implicit bias. *Underst. Interv.* 9(2). **https://www.understandinginterventionsjournal.org/article/6355-breaking-the-bias-habit-a-workshop-to-help-internal-medicine-residents-reduce-the-impact-of-implicit-bias**
Pedersen A, Barlow FK. 2008. Theory to social action: a university-based strategy targeting prejudice against Aboriginal Australians. *Aust. Psychol.* 43(3):148–59
Pendry LF, Driscoll DM, Field SCT. 2007. Diversity training: putting theory into practice. *J. Occup. Organ. Psychol.* 80(1):27–50
Pettigrew TF, Tropp LR. 2006. A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* 90(5):751–83
Pettijohn TF II, Walzer AS. 2008. Reducing racism, sexism, and homophobia in college students by completing a psychology of prejudice course. *Coll. Stud. J.* 42(2):459–66
Phillips BN, Deiches J, Morrison B, Chan F, Bezyak JL. 2016. Disability diversity training in the workplace: systematic review and future directions. *J. Occup. Rehabil.* 26(3):264–75

Prescott-Clements L, Schuwirth L, van der Vleuten C, Hurst Y, Whelan G, et al. 2013. The cultural competence of health care professionals: conceptual analysis using the results from a national pilot study of training and assessment. *Eval. Health Prof.* 36(2):191–203

Preusser MK, Bartels LK, Nordstrom CR. 2011. Sexual harassment training: person versus machine. *Public Pers. Manag.* 40(1):47–62

Rehg MT, Gundlach MJ, Grigorian RA. 2012. Examining the influence of cross-cultural training on cultural intelligence and specific self-efficacy. *Cross Cult. Manag.* 19(2):215–32

Renzaho AMN, Romios P, Crock C, Sonderlund AL. 2013. The effectiveness of cultural competence programs in ethnic minority patient-centered health care—a systematic review of the literature. *Int. J. Qual. Health Care* 25(3):261–69

Reynolds L. 2010. Aging and disability awareness training for drivers of a metropolitan taxi company. *Act. Adapt. Aging* 34(1):17–29

Ridley CR, Baker DM, Hill CL. 2001. Critical issues concerning cultural competence. *Couns. Psychol.* 29(6):822–32

Roberson L, Kulik CT, Pepper MB. 2001. Designing effective diversity training: influence of group composition and trainee experience. *J. Organ. Behav.* 22(8):871–85

Roberson L, Kulik CT, Pepper MB. 2009. Individual and environmental factors influencing the use of transfer strategies after diversity training. *Group Organ. Manag.* 34(1):67–89

Roberts SO, Bareket-Shavit C, Dollins FA, Goldie PD, Mortenson E. 2020. Racial inequality in psychological research: trends of the past and recommendations for the future. *Perspect. Psychol. Sci.* 15(6):1295–309

Rogers-Sirin L, Sirin SR. 2009. Cultural competence as an ethical requirement: introducing a new educational model. *Divers. High. Educ.* 2(1):19–29

Ross SN. 2014. Diversity and intergroup contact in higher education: exploring possibilities for democratization through social justice education. *Teach. High. Educ.* 19(8):870–81

Rudman LA, Ashmore RD, Gary ML. 2001. "Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes. *J. Pers. Soc. Psychol.* 81(5):856–68

Rynes S, Rosen B. 1995. A field survey of factors affecting the adoption and perceived success of diversity training. *Pers. Psychol.* 48(2):247–70

Saha S, Korthuis PT, Cohn JA, Sharp VL, Moore RD, Beach MC. 2013. Primary care provider cultural competence and racial disparities in HIV care and outcomes. *J. Gen. Intern. Med.* 28(5):622–29

Sanchez JI, Medkik N. 2004. The effects of diversity awareness training on differential treatment. *Group Organ. Manag.* 29(4):517–36

Sanchez-Burks J, Lee F, Nisbett R, Ybarra O. 2007. Cultural training based on a theory of relational ideology. *Basic Appl. Soc. Psychol.* 29(3):257–68

Schellhaas FMH, Dovidio JF. 2016. Improving intergroup relations. *Curr. Opin. Psychol.* 11:10–14

Schim SM, Doorenbos AZ, Borse NN. 2006. Enhancing cultural competence among hospice staff. *Am J. Hosp. Palliat. Care* 23(5):404–11

Schmidt CK, Earnest DR, Miles JR. 2019. Expanding the reach of intergroup dialogue: a quasi-experimental study of two teaching methods for undergraduate multicultural course. *J. Divers. High. Educ.* 18(3):264–73

Stake JE, Hoffmann FL. 2001. Changes in student social attitudes, activism, and personal confidence in higher education: the role of women's studies. *Am. Educ. Res. J.* 38(2):411–36

Stewart E. 2018. Starbucks says everyone's a customer after Philadelphia bias incident. *VOX*, May 19

Stone J, Moskowitz GB. 2011. Non-conscious bias in medical decision making: What can be done to reduce it? *Med. Educ.* 45(8):768–76

Sue DW. 2001. Multidimensional facets of cultural competence. *Couns. Psychol.* 29(6):790–821

Tankard ME, Paluck EL. 2016. Norm perception as a vehicle for social change. *Soc. Issues Policy Rev.* 10(1):181–211

Thom DH, Tirado MD, Woon TL, McBride MR. 2006. Development and evaluation of a cultural competency training curriculum. *Med. Educ.* 6(1):6–38

Thomas VJ, Cohn T. 2006. Communication skills and cultural awareness courses for healthcare professionals who care for patients with sickle cell disease. *J. Adv. Nurs.* 53(4):480–88

VanderWeele TJ. 2013. Surrogate measures and consistent surrogates. *Biometrics* 69(3):561–69
Vianden J. 2018. "In all honesty, you don't learn much": White college men's perceptions of diversity courses and instructors. *Int. J. Teach. Learn. High. Educ.* 30(3):464–76
Waight J, Madera JM. 2011. Diversity training: examining minority employees' organizational attitudes. *World Hosp. Tour. Themes* 3(4):365–76
Walton GM, Cohen GL. 2007. A question of belonging: race, social fit, and achievement. *J. Pers. Soc. Psychol.* 92(1):82–96
Weber SJ, Cook TD. 1972. Subject effects in laboratory research: an examination of subject roles, demand characteristics, and valid inference. *Psychol. Bull.* 77(4):273–95
Weech-Maldonado R, Elliott MN, Pradhan R, Schiller C, Hall A, Hays RD. 2012. Can hospital cultural competency reduce disparities in patient experiences with care? *Med. Care* 50:S48–55
Wicker AW. 1969. Attitudes versus actions: the relationship of verbal behavioral responses to attitude objects. *J. Soc. Issues* 25:41–78
Xu Y, Shen J, Bolstad AL, Covelli M, Torpey M. 2010. Evaluation of an intervention on socio-cultural communication skills of international nurses. *Nurs. Econ.* 28(6):386–92
Yoder JD, Mills AS, Raffa ER. 2016. An effective intervention in research methods that reduces psychology majors' sexist prejudices. *Teach. Psychol.* 43(3):187–96
Zúñiga X, Naagda BA, Sevig TD. 2002. Intergroup dialogues: an educational model for cultivating engagement across differences. *Equity Excell. Educ.* 35(1):7–17

Annual Review of Psychology

Volume 73, 2022

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at http://www.annualreviews.org/errata/psych