Immersive Inclusion: Diversity and Inclusion Training Using Virtual Reality

A thesis submitted to the

Graduate School

of the University of Cincinnati

in partial fulfillment of the

requirements for the degree of

Master of Arts

in the Department of Psychology

of the College of Arts and Sciences

by

Lauren M. Mason

B.A. Randolph College

May 2016

M.A. University of Cincinnati

April 2019

Committee Chair: Donna Chrobot-Mason, Ph.D.

Abstract

Diversity, equity, and inclusion (DEI) training is widely used by companies, but often with limited or unknown effectiveness. DEI programs incorporating immersive virtual reality (VR) have shown promising results due to a unique impact on empathy and behavioral intention. The purpose of the current research is to add to the literature on VR as a DEI training tool and investigate the lasting effects of a VR-based DEI training program. Study 1 was longitudinal and investigated the impacts of a VR-based bias training. Cognitive and affective empathy levels and impact on behavior, attitude, and knowledge before and after the training were measured to test the hypotheses that (H1) cognitive empathy levels would increase and (H2) individuals with higher initial levels of empathy would demonstrate more pronounced changes in cognitive empathy following the training. H1 was supported but larger changes were found in affective empathy levels. H2 was also supported as individuals with higher initial empathy levels showed higher levels of cognitive empathy after the training compared to individuals with lower initial empathy levels. However, again, larger differences were found in affective empathy levels. Qualitative data revealed a lasting impact nine weeks after the training that was not present in the quantitative data. Study 2 surveyed healthcare professionals who previously participated in a VR-based DEI training that focused on social determinants of health and empathy in healthcare professionals. The purpose of this study was to gain insight into the longitudinal impacts of a VR-based DEI training by gathering qualitative data from the participants at least a year after they went through the training. The respondents reported a lasting influence from the training. Reasons for the discrepancy between the qualitative and quantitative results are discussed as are implications for organizations and future DEI training development and research.

Keywords: diversity and inclusion training, virtual reality, empathy, perspective taking

Acknowledgements

Although I am listed as the sole author of this manuscript, there are many others without whom this thesis would not exist. I'd like to start by expressing my sincerest gratitude to my thesis chair and mentor, Dr. Donna Chrobot-Mason, for seeing something in me that even I couldn't see yet. Without you I wouldn't be where I am today and I'm truly grateful for your continued guidance and patience.

I also want to thank Dr. Stacie Furst-Holloway for providing invaluable insight and expertise throughout this journey. Dr. Susan Brammer, thank you for collaborating with me and allowing me access to an entirely new population. A very special thank you to Myra Laldin for sticking with me throughout this process. This project truly could not have happened without you and your brilliant team at Perspectives.

Thank you to my coding team – Erin Barta, Kristen Fisher, Kunal Kewalramani, and Ben Marthey – for bearing with me through long discussions. I also have to thank my dear friends and PhD hopefuls – you know who you are – for being my support system and sounding board. Finally, thank you to my parents, my sister, and Mevan whose shoulders I can always lean on. This would not be possible without the love and support of all of the amazing people in my life.

Table of Contents

List of Tables and Figures

Immersive Inclusion: Diversity and Inclusion Training Using Virtual Reality

Diversity and inclusion in the workplace has had a complex and controversial history since its start in the 1960s (Kelly & Dobbin, 1998). Through the decades, diversity initiatives have shifted focus multiple times in response to public pressures, which has not always produced the most effective programs (Anand & Winters, 2008). Recently, following the national and international Black Lives Matter protests of 2020, many well-known companies have made commitments to accelerate their diversity efforts (Westerman et al., 2020). Among others, Starbucks, General Motors, and Walmart made pledges to increase the diversity of their workforces, make anti-bias training mandatory, and even link executive pay to the achievement of diversity efforts (Gurchiek, 2020). There are various reasons that companies are making these renewed efforts. Companies are facing pressures from activists among their own employees and from external sources like the Nasdaq, a national securities exchange, that are pushing for more diversity on company boards (Sorkin et al., 2020).

In addition to internal and external pressures, there is a myriad of research-backed benefits that incentivize businesses to increase their diversity and inclusion efforts. Businesses with higher diversity have lower turnover rates and higher productivity than businesses with low levels of diversity (Catalyst, 2020). Employees are more motivated and engaged when they feel included at work and supported by their company (Luu et al., 2019). Inclusion efforts improve a company's reputation, making them more appealing to potential employees and consumers because they are seen as more ethical and responsible (Rabl et al., 2018; Zalis, 2019).

Despite all of these proven benefits of diversity and inclusion in the workplace, most programs meant to improve diversity and inclusion have been shown to be ineffective at best and cause more diversity-related issues at worst (Chang et al., 2019). A major reason that diversity

programs fail is that they are implemented poorly and end up making participants defensive and more resistant to changing their behavior (Asare, 2019; Duguid & Thomas-Hunt, 2015). Additionally, there is a lack of evaluation and empirical research on most diversity training programs. More research is needed to determine the causal effects of diversity training methods and which methods are the most effective. Some newer diversity programs have started using virtual reality to create more nuanced and engaging trainings with promising results (Bertrand et al., 2018; Herrera et al., 2018; van Loon et al., 2018). However, more research is needed on virtual reality as a training tool to develop cognitive and emotional skills like the ones targeted by diversity training programs.

The current study investigated virtual reality as a training tool for diversity and inclusion programs and explored the effects of these types of trainings. In this paper, I will first discuss the current state of the literature surrounding diversity and inclusion trainings, including what methods are and are not effective. Second, I will explain why empathy development is a key factor for instigating behavioral changes. Third, I will unpack the research surrounding virtual reality and explain why it is well-suited for diversity training. Finally, I will explain how the current study contributes to the literature and why the study took place across two data collection sites.

Diversity Training Successes and Failures

As a result of a number of high-profile lawsuits in the early 2000s, almost half of midsized organizations and nearly all Fortune 500 companies offer some type of diversity, equity, and inclusion (DEI) training to their employees (Chang et al., 2019; Dobbin & Kalev, 2016). Despite the large investment in and widespread implementation of diversity training, traditional methods of DEI trainings have been shown to have limited effectiveness, especially in

the long-term (Bezrukova et al., 2016; Bezrukova et al., 2012; Paluck & Green, 2009). Additionally, studies show that many diversity trainings can cause adverse effects in the companies that implement them (Duguid & Thomas-Hunt, 2015; Noon, 2018). This is due to a variety of reasons.

Many companies only use diversity training as remedial – required training in response to an issue that often targets just one group of employees that they have had problems with (Dobbin & Kalev, 2016). Using diversity training in this way makes it a punishment rather than an opportunity to foster an inclusive workplace environment and often makes individuals defensive instead of open to change. While many existing DEI trainings are focused on one aspect of diversity and inclusion, such as compliance issues or stereotypes, research has shown that an integrated approach is more effective (Bezrukova et al., 2012). An integrated approach takes more time and effort because it involves having much broader goals for the training, such as understanding systemic issues, and making DEI goals part of an organization's development plan (Bendick et al., 2001).

The content and delivery of a DEI training has the potential to either increase or decrease instances of bias within a workplace. There is evidence to suggest that DEI trainings that focus on the high prevalence of stereotyping within the general population can result in increasing the expression of stereotyping by participants (Duguid & Thomas-Hunt, 2015; Noon, 2018). Duguid & Thomas-Hunt (2015) conducted multiple studies on how diversity trainings can actually condone stereotypes. They found that individuals exposed to messages about the high prevalence of stereotyping in the general population were more likely to express stereotype bias compared to those exposed to messages about the low prevalence of stereotyping or the high prevalence of counterstereotyping (the effort to avoid stereotyping; Duguid & Thomas-Hunt, 2015). These

findings indicate that it is not enough to simply point out the issues of stereotyping and bias within workplaces or the general population, rather DEI trainings need to go a step further and teach individuals how to counteract stereotyping and bias to foster inclusion.

Pointing out what individuals are doing wrong without teaching them what they should be doing will usually make individuals defensive and unwilling to engage with the training material (Asare, 2019; Duguid & Thomas-Hunt, 2015). Diversity and inclusion trainings should aim to help individuals overcome their implicit biases and prejudices so that they will be more inclined to engage with people outside of their "in-group" (Shore et al., 2018). Studies show that individuals are more likely to do this when DEI training incorporates strategies that strengthen their diversity self-efficacy (DSE; Combs, 2002; Combs & Luthans, 2007). Combs and Luthans (2007) define DSE as the belief that one has the necessary motivation and resources to change behavior and realize diversity goals in the workplace. DSE is an essential component to setting up lasting intentions for behavior change following a DEI training, especially when training leaders to navigate and manage an increasingly diverse workforce (Combs, 2002). Setting goals or intentions has been shown to be one of the more effective approaches to DEI training along with perspective taking (Galinsky & Moskowitz, 2000; Lindsey et al., 2015; Lindsey et al., 2017).

Success of Perspective Taking Exercises and Their Relationship to Empathy

Perspective taking is a cognitive exercise that involves considering situations from the viewpoint of another (Galinsky & Moskowitz, 2000). Research has shown that, when used in DEI training, perspective taking exercises result in attitudinal changes that are linked to an increase in cognitive empathy (Batson et al., 1996; Lamm et al., 2007; Vescio et al., 2003).

Being able to empathize with another person or group, especially someone outside of your own in-group, is vital to promoting progress toward inclusion.

Research has shown that individuals with a higher level of empathy are more likely to exhibit altruistic behavior than those with a lower level of empathy (Bethlehem et al., 2016; Krebs, 1975; Smith, 2008). A possible reason for this is that feeling the distress of another person serves as motivation to reduce one's own distress by helping the other person (de Waal, 2008). Empathy is also a strong motivator for justice and a predictor for higher sensitivity to injustice for others (Decety & Yoder, 2015). These conclusions point to individual empathy levels as strong indicators of motivation to act when injustice in the workplace is observed.

Empathy has been recognized as being a multidimensional construct with four main components: perspective taking, taking on the viewpoint of others; fantasy, inserting oneself into the feelings and actions of fictional characters; empathic concern, feeling sympathy or concern for others; and personal distress, feeling anxious or uneasy in tense interpersonal settings (Davis, 1983). Two of these components (perspective taking and fantasy) require cognitive abilities to imagine what another person is feeling or how they might perceive a situation. The other two components (personal distress and empathic concern) require affective abilities to feel something on behalf of another or in response to a situation. Because empathy has both cognitive and affective aspects, the method used to train empathy should allow and encourage the participant to experience both aspects fully.

Within recent decades, companies have begun to recognize the importance of interpersonal skills, like empathy, for leadership and have implemented training programs targeting empathy-related skills (Doe et al., 2015). Most empathy-related skill development programs incorporate a mix of lecture, role-play, and discussion delivery methods for teaching

the material (Mattingly & Kraiger, 2019; Sogunro, 2004). Participants receive feedback from the instructor or their fellow peers in the program (Sogunro, 2004). Research has shown these workplace training programs to be moderately effective at teaching empathy-related skills (Arthur et al., 2003; Groves et al., 2006; Mattingly & Kraiger, 2019). However, there is no research on the long-term effects of these training programs. Skill levels may increase immediately following the training program, but those scores could differ a few weeks later, especially if employees do not have the opportunity in their jobs to practice the skills (Arthur et al., 2003).

Most perspective taking exercises attempt to increase empathy by asking participants to engage in imagined situations (Galinsky & Moskowitz, 2000; Shih et al., 2009; Vescio et al., 2003). These trainings are limited because they rely on participants' ability to imagine a situation from another's perspective. In other words, they must already be able to empathize with that population. Participants who would most benefit from diversity trainings are individuals who have not had much experience with populations outside of theirs, meaning that they would not get much out of most diversity trainings if prior knowledge of other cultural perspectives is required.

One diversity training program in the hospitality industry went a step further by incorporating a behavioral manipulation that asked participants to complete a recipe using only nonverbal communication to simulate that they could not speak or understand English (Madera et al., 2011). Madera and others (2011) had more success than traditional DEI trainings in increasing participants' empathy for and understanding of the experience of non-English speakers in the hospitality industry, but this training could be improved even further by using immersive virtual reality.

Immersive Virtual Reality as a Training Tool

Immersive virtual reality (VR) is a computer generated, three-dimensional environment that replicates the basic features of reality by giving the user a first-person perspective. The user views the virtual world through a pair of three-dimensional goggles, or a head-mounted display, giving the user a high degree of immersion because they are visually surrounded by the training simulation (Kapp & O'Driscoll, 2010). With the addition of headphones and hand-held controllers, the user can fully replace perceptual stimulation from the real world with that of the virtual environment. This is what allows users to fully experience someone else's perspective (De la Peña et al., 2010).

VR has been shown to be an effective perspective taking tool in several studies (Bertrand et al., 2018; Herrera et al., 2018; van Loon et al., 2018). Perspective taking trainings using VR not only demonstrate lasting attitudinal changes, but one study found additional behavioral changes not present in the outcomes of traditional perspective taking exercises (Herrera et al., 2018). Herrera and others (2018) found lasting attitude and behavioral changes toward homeless individuals following their VR simulation called "Becoming Homeless" in which participants experience what it is like to lose their jobs, get evicted, and spend the night on a bus. They found that the changes lasted for at least two months after the experience (Herrera et al., 2018). Virtual reality has the potential for additional training benefits due to the immersive nature of the VR experience (Bertrand et al., 2018; Schmid Mast et al., 2018). VR adds a first-person immediacy to a perspective taking exercise that makes the trainee an active participant in the scenario so that the events are happening to them rather than to someone they are observing (Kapp & O'Driscoll, 2010). This feeling of presence within the VR environment is one of the key components of a VR training that results in greater knowledge retention.

Impact of Presence and Agency within VR

The first-person perspective combined with the perceptual immersion of VR creates an embodied experience or a feeling of presence for the participant, meaning that they react to an event as if it is happening to them in real life (Lee, 2004). For example, if the VR scenario puts them in a tense interpersonal situation, participants may react with slight perspiration or increased heart rate. This emotional and physiological reaction leads to greater retention of learned information and increased impact on behavioral intention (Makransky & Lilleholt, 2018).

Trainings within VR involve a constructivist approach to learning by giving the participant agency, or the ability to make choices within the experience (Karagiorgi & Symeou, 2005). Allowing the participant to be involved in the progression of the training scenario means that they can directly address their own perceptions and biases by bringing them into the learning experience, adding an aspect of self-reflexivity that results in deeper learning (Bertrand et al., 2018). This aspect of co-creation of the learning experience means that participants can lead with their own biases to engage in reflexivity by making mistakes and finding their own answers (Karagiorgi & Symeou, 2005).

Role of Cognitive Load in Perspective Taking

Additionally, VR is particularly suited for perspective taking exercises because experiencing another person's perspective within VR requires no cognitive work on the part of the participant. Traditional perspective taking exercises usually require the participant to read a scenario and imagine the perspective of another person. Imagining another's perspective requires considerable cognitive resources and can limit a participant's available working memory while they are engaging in perspective taking (Cane et al., 2017; Rossnagel, 2000). Because of the

immersion inherent in a VR environment, this means that a participant does not need to imagine a perspective, they simply experience it. This leaves their cognitive resources available for empathizing with that person and processing their experience.

Because of the cognitive load, traditional perspective taking exercises require a lot of intrinsic motivation from the participant to get the most out of the exercise (Cane et al., 2017). Research has shown that individuals' motivation to engage in a training is higher when they do not have to work as hard, especially for those individuals who do not already recognize the importance of DEI trainings (Holladay et al., 2003; Kulik et al., 2007).

Specificity of Context and Transfer of Training into the Workplace

In the past two decades, VR has been used to train technical skills and emergency preparedness in the medical field, industrial maintenance, and even the Army (Gallagher et al., 2005; Gavish et al., 2015; Hsu et al., 2013; Wyld, 2010). VR is well suited for these types of trainings because the situations feel real and can be tailored to include specific circumstances, which enhances the transfer of training into the workplace (Cox et al., 2017; Schmid Mast et al., 2018).

Several branches of the Army use virtual platforms, including VR, to train interpersonal skills and prepare officers to work in various cultural contexts before going overseas (Campbell et al., 2011; Deaton et al., 2005; Hays et al., 2012). The most effective interpersonal skill training programs are active and personal, meaning that participants have multiple opportunities to practice the skills and receive plenty of individual feedback (Mattingly & Kraiger, 2019). A VR environment allows a safe space to practice new skills and experience realistic consequences without any real-life consequences. The participant receives immediate feedback on their choices from the avatars in the virtual environment (Schmid Mast et al., 2018). This type of skill building

directly translates to real-life application in a safe and efficient way. Because of the relevancy and realism of a VR-based training, participants are more likely to experience lasting behavioral changes compared to traditional training methods (Herrera et al., 2018; Schmid Mast et al., 2018).

Individual Differences in Response to Diversity Training

The prior knowledge and experience that individuals bring to a DEI training is an important consideration that is often overlooked. There is evidence to suggest that individuals with close cross-race friendships are more motivated to participate in DEI courses and recognize the importance of DEI for their career than those without close cross-race friendships (Ragins & Ehrhardt, 2020). This friendship effect resulted in a larger improvement in perspective taking skills among those with close cross-race friendships compared to those without (Ragins & Ehrhardt, 2020). Other studies have found that individuals' baseline empathy levels can impact the effects of perspective taking training (Lindsey et al., 2015; Meadows et al., 2017). A study on anti-fat bias in the medical community found that the training was more effective on individuals who were more empathetic at baseline (Meadows et al., 2017). Another study on attitudes toward gay and African American individuals found that the perspective taking training was more effective on participants with low baseline empathy (Lindsey et al., 2015). These findings indicate that it is important to track baseline empathy alongside training impact to ensure that diversity training can be as effective as possible for all participants.

The Current Study

VR is still new to the DEI space and, while there is some research to support VR as a learning tool, there is still much that is not known. The current study adds to the literature by contributing findings on the impact of VR-based DEI training on emotion, knowledge, and

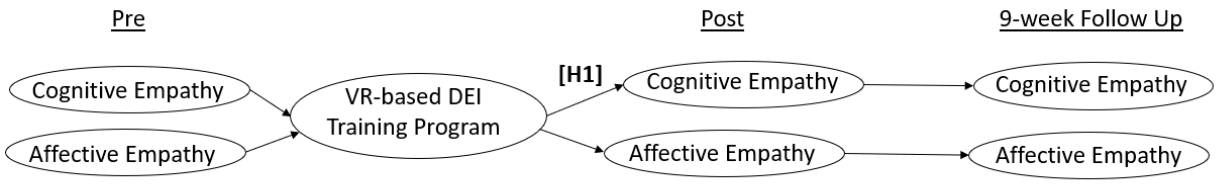
behavior in participants. This information is necessary to improve existing DEI trainings and create more impactful trainings in the future.

To investigate this topic, I conducted two studies. For the primary study, I partnered with Perspectives, a company in Cincinnati that offers immersive leadership inclusion training using virtual reality. Their training is based on creating empathy, recognizing exclusive behaviors, responding as an ally, and realizing the impact these behaviors have on an organization as a whole. Within the study, I differentiated between cognitive and affective empathy because research has shown that cognitive empathy is likely the trainable aspect of empathy and affective empathy is likely the more stable personality trait (Barlinska et al., 2018; Rueda et al., 2015; van Loon et al., 2018; Zahn-Waxler et al., 1992). Based on this, my first hypothesis was that cognitive empathy (not affective empathy) skills would be higher after the training compared to before. Figure 1 illustrates the model for this first hypothesis.

H1: Cognitive empathy (not affective empathy) skills will be higher after the training compared to before.

Figure 1

Model for Hypothesis 1



Note: H1: **Cognitive empathy** (not affective empathy) skills will be higher after the training compared to before.

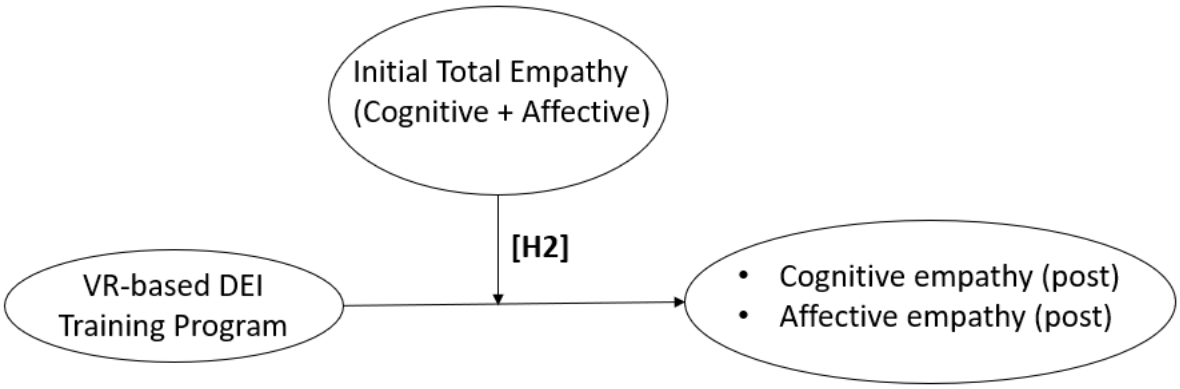
Following that same line of reasoning, my second hypothesis was that individuals' initial level of empathy would moderate the impact of the training on post-training empathy levels so that individuals with higher levels of empathy, both affective and cognitive, would demonstrate a

more pronounced increase in post-training cognitive empathy levels (not affective empathy levels) than those with lower initial empathy levels. Figure 2 illustrates the model for the second hypothesis.

> *H2: Individuals with higher initial empathy levels (both cognitive and affective) will demonstrate a larger increase in cognitive empathy (not affective empathy) after training compared to individuals with lower initial empathy levels.*

Figure 2

Model for Hypothesis 2



Note: H2: Individuals with higher **initial empathy** (both affective and cognitive) will demonstrate a higher increase in **cognitive empathy (not affective empathy) after training** than those with lower initial empathy levels.

For the second study, I partnered with the University of Cincinnati College of Nursing (UCCON). They have an ongoing grant with the Medicaid Technical Assistance and Policy Program (MEDTAPP) that allowed them to create a virtual reality simulation depicting patients challenged by social determinants of health to be used as a clinical tool to train providers caring for these populations. Social determinants of health are "the conditions in the environments

where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks," which may include racism, discrimination, education, income, language skills, etc. (U.S. Department of Health and Human Services, n.d.). The goal of the project was to raise awareness of and empathy for these issues in local health care providers. My study surveyed previous participants of the simulation training to gather qualitative data from a different context than the primary study to further explore diversity and inclusion training using VR. Because at least a year passed for these participants, this data provided insight into the longitudinal impacts of an empathy-focused VR-based training. This study was meant to explore the impact of the training on the participating healthcare providers. While there was not any baseline data to compare to from before the training, the qualitative data from the participants was still valuable for gaining insight into how they felt the training may have impacted them and their work.

Together the studies were meant to provide evidence of the impact of a DEI training that uses VR. The first study focused more on empathy and behavioral intention immediately following a training and had the additional component of comparing data at three timestamps: before the training, immediately after the training, and nine weeks after the training. The second study added to the first by providing more insight on the longitudinal impact of a training as it had been at least a year since any of the healthcare professionals participated in the simulation training. The longer time period between the training and follow up in the second study gave participants more opportunity to apply what they learned from the training in their workplaces. This resulted in rich qualitative data about whether the participants were able to apply anything from the training into their work and specific examples of application.

The VR simulations in each study approached bias in different ways. The simulations developed by Perspectives in Study 1 depict instances of workplace bias in the form of microaggressions against an individual. The simulations developed the MEDTAPP grant in Study 2 allowed healthcare professionals to see into patients' lives outside of their medical appointments and experience possible barriers that they might be dealing with when trying to access healthcare. This allowed the healthcare professionals to recognize their own biases against these patients when they cannot show up to appointments or follow their plan of care. The training in Study 1 was focused on teaching participants to recognize and respond to situations of bias, while the training in Study 2 focused on creating an understanding of the systemic biases that could be contributing to a patient's behavior. By investigating the impacts of these differing approaches to bias training, the current study provides multiple avenues of evidence for what the impacts of a VR-based DEI training may be and how training impact occurs.

Study 1: Perspectives

Method

Participants

The participants all worked within the same engineering department within a mid-sized, for-profit company in the mid-west. The department hired Perspectives (the VR training company) for their inclusion training package and asked all the employees in the department to go through it as they were able. Approximately 60 employees participated in the training program. In total, 58 employees responded to the study surveys and 45 of them answered the demographic questions. Of those employees, 29 were male (nine female and seven prefer not to answer), 34 were White (five non-white and six prefer not to answer), and the average age was 35-44 years old.

Materials

Participants were surveyed four times throughout the training: pre-survey (one-week before the training), VR survey (after going through the VR scenario(s)), post-survey (after last training session), follow-up survey (nine-weeks after the last training session). All surveys were administered through Typeform, an online survey platform. The link for each survey was sent to the participants through their work email. The pre-survey, post-survey, and the follow-up survey included the same questions regarding participants' cognitive and affective empathy levels. Table 1 provides information on the four surveys administered to collect data in the study including the names used to refer to each survey, the measures on each one, and when they were administered within the training timeline.

Table 1

Breakdown of the Four Surveys Included in the Study, the Measures on Each, and When they Were Administered Within the Training Timeline

Survey Name	Measures	Timeline
Pre Survey	cognitive empathy, affective empathy	administered one week before the first training session
VR Survey	immersion, feelings as Max	administered after the first (Group 1) or only (Group 2) VR training session
Post Survey	cognitive empathy, affective empathy	administered after the last training session
Follow-up Survey	cognitive empathy, affective empathy	administered nine-weeks after the last training session

The VR survey (the survey administered after participants experienced the VR scenario(s)) asked about the participants' feeling of immersion within the scenario and their feelings as Max (the character from whose perspective the participants experienced the VR scenario) during the experience. All four surveys (pre, post, follow-up, and the VR survey) also

included open-ended questions to better capture participants' knowledge, reactions, and thoughts throughout the training. All of the survey items included in the study can be found in Appendix A.

The virtual reality experiences were developed prior to the study by Perspectives. More information about their development process can be found on their website (Perspectives, 2020). Perspectives has developed multiple VR experiences that focus on different experiences of bias, each with an accompanying inclusion version of the scenario demonstrating how to act as an ally. The client company for this study wanted to focus on racial bias for their training. Perspectives offered them their "Max Experience" – a scenario experienced from the perspective of Max, an African American account manager, who is meeting with a new client, Patrick, along with his coworker, Chad, who is also an account manager. In the exclusion version of the scenario, various microaggressions are committed by Patrick and Chad towards Max. In the inclusion version of the scenario, Patrick behaves the same but Chad acts as an ally for Max. The participants used Oculus headsets provided by Perspectives to go through the VR scenarios.

Measures

Two subscales from the Interpersonal Reactivity Index (IRI) were used as the empathy measure (Davis, 1980). The two 7-item Likert-style subscales were perspective taking, which measures the tendency to take on the point of view of other people (this served as the measure of cognitive empathy), and empathic concern, which measures the tendency to feel compassion and concern for others in negative situations (this served as the measure of affective empathy; Davis, 1980). Sample items from the perspective taking and empathic concern subscales respectively are: "Before criticizing somebody, I try to imagine how I would feel if I were in their place."; "I often have tender, concerned feelings for people less fortunate than me." In the original study

developing the IRI, the coefficient alphas for each subscale were split between males and females and they ranged from 0.68 to 0.75 for the empathic concern and perspective taking subscales (Davis, 1980). The items were answered on a scale of 1 (not at all like me) to 5 (very much like me). Scores were obtained by totaling the seven items, meaning the possible score range was 7-35.

Design and Procedure

The participants went through the training as two separate groups. Group 1 completed the training over six weeks in March – April 2021. Group 2 completed the training over five weeks in May – June 2021. The content of the training was the same across both groups with small changes in the timing and delivery, which are explained in the following paragraphs.

During the training weeks, the groups met once a week for a training session, usually for an hour and a half. The first session was a brief introduction meeting letting participants know what to expect from the training. The pre survey was sent out after the first meeting and the actual training started in the second week. During the second week, Group 1 went through the exclusion version of the VR scenario and completed a worksheet asking them to write down the exclusive behaviors that they noticed during the experience. The VR survey was sent out after this session. Group 1 discussed the exclusion scenario in the third week while also discussing behaviors they had noticed at their workplace that they felt needed to be addressed to create a more inclusive culture. Group 1 went through the inclusion version of the VR scenario in the fourth week and completed a worksheet asking them to list any inclusive behaviors that they noticed during the scenario. Group 1 discussed the inclusion scenario during the fifth week while also learning how to spot or recognize exclusive behaviors, call attention to them without calling out individuals (“call-in”), and give productive feedback about how to move forward.

Group 2 went through the same activities as Group 1 except that Group 2 went through both VR scenarios in one two-hour session, completing the worksheets and engaging in a 30-minute discussion after each scenario. Thus, Group 1's training schedule was more spread out while Group 2's was more condensed. In the final session, the participants created their own action plan to integrate their learnings into their existing culture. The post-survey was sent out right after the final session. Nine weeks after the final session, the follow-up survey was sent out. The full training schedules for both participants groups can be seen in Table 2. All the training sessions were held on Microsoft Teams except for the VR sessions, which were held in person at a training facility. During the sessions, the participants were split into small groups of 5-7 people and spent most of the session completing activities within those small groups and came back together as a large group at the end to share. All of the training sessions were facilitated by a Perspectives team member.

Table 2

Training Schedule

	Participant Group 1 (spread out)	**Participant Group 2 (condensed)**
Week 1	**Kick off** *[pre-survey]*	**Kick off** *[pre-survey]*
Week 2	**VR Session 1:** Max Experience Exclusion *[VR-survey]*	**VR Session:** Max Experience Exclusion & Inclusion with discussions *[VR-survey]*
Week 3	**Conversation 1:** Exclusion VR discussion & Behaviors to address in our workplace	**Conversation 1:** Behaviors to address in our workplace
Week 4	**VR Session 2:** Max Experience Inclusion	**Conversation 2:** Spot, Call-in, Feedback
Week 5	**Conversation 2:**	**Conversation 3:** Commitment to Action

	Inclusion VR discussion & Spot, Call-in, Feedback	*[post-survey]* *[follow-up survey nine weeks later]*
Week 6	**Conversation 3:** Commitment to Action *[post-survey]* *[follow-up survey nine weeks later]*	

Note. The participant groups were led through the same activities and discussions but experienced and discussed the VR scenarios on differing timelines. "Spot, Call-in, Feedback" refers to a specific training lesson in which participants learned how to spot or recognize exclusive behaviors, call attention to them without calling out individuals ("call-in"), and give productive feedback about how to move forward.

Results

Quantitative

All analyses were performed using SPSS. Table 3 includes the total response counts for each survey and response group. The data was tested for normality using the Shapiro-Wilk test and was found to be normally distributed. There were no significant outliers found within the data. The assumption of sphericity was found to be violated by at least one variable within the dataset, which meant that a repeated-measures analysis of variance (ANOVA) was not an appropriate fit for the data (Krueger & Tian, 2004). To assess whether it was appropriate to combine the two participant groups for purposes of analyses, chi-square tests of independence were conducted with each of the demographic variables. There was no association found between gender ($G^2(2, N = 39) = 0.28$, $p = 0.903$), race ($G^2(4, N = 39) = 3.54$, $p = 0.472$), or age ($G^2(4, N = 40) = 1.02$, $p = 0.907$) and participant group. The likelihood ratio statistic was reported for all three tests because some of the categories had less than 10 counts in them. Based on these tests, the groups were combined for most of the analyses.

Table 3

Response Counts for Each Survey and Participant Group

	Group 1	Group 2	Total
Pre Survey	25	16	41
VR Survey	23	21	44
Post Survey	17	11	28
Follow-up Survey	11	7	18
Total	27	31	58

Because of the small sample sizes, a confirmatory factor analysis (CFA) could not be conducted to further explore the relationship between the variables used in the study. A power analysis conducted using G*Power version 3.1.9.7 (Faul et al., 2007) revealed that a sample size of $N = 400$ would be needed to conduct a CFA to achieve 80% power with a medium effect size at alpha level 0.05. It is recommended to have a sample size of at least 100 to achieve suitable power in a factor analysis (Costello & Osborne, 2005; Wolf et al., 2013). Thus, it was determined that linear regression would be a more appropriate statistical analysis to use given the small sample size. The empathic concern and perspective taking subscales had a moderate correlation ($r = 0.41$, $p = 0.05$).

The first hypothesis (H1) proposed that cognitive empathy levels (not affective empathy levels) would be higher after the training compared to before. A linear mixed model analysis (LMM) was conducted to test this hypothesis. Using this analysis, the variables could be looked at longitudinally across the three time points (pre, post, and follow-up) and across the two participant groups. A LMM was chosen instead of a repeated-measures ANOVA because a LMM allows for incomplete data among the participants (Fitzmaurice & Ravichandran, 2008; Krueger & Tian, 2004; Misangyi et al., 2006; Quené & Van Den Bergh, 2004). To further

explain, there were a fair amount of participants in the current study who completed one or two of the three surveys (pre-, post-, or follow-up) but not all three. A repeated-measures ANOVA would remove any participant who did not complete any one of the three surveys, which would reduce power and sample size. A LMM allows participants with missing data to remain in the data set, which increases the sample size and the power for the study.

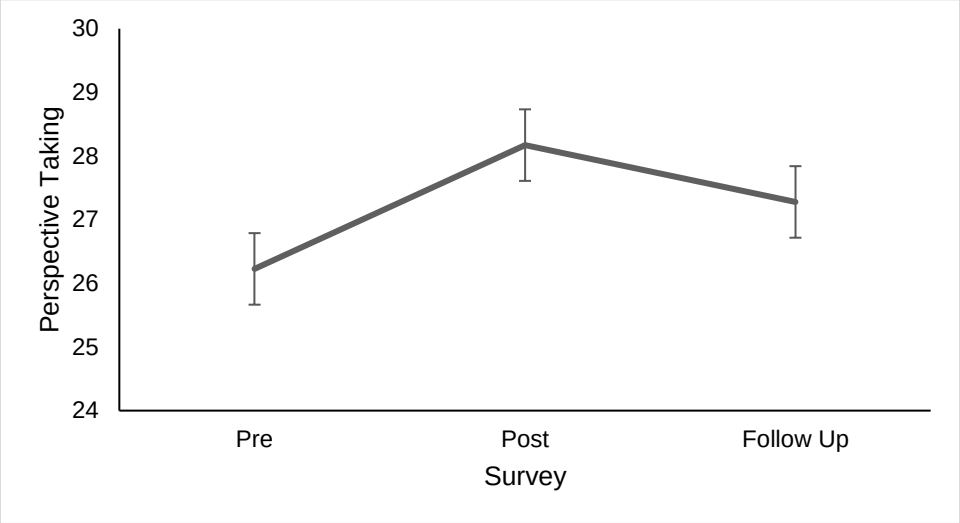
The measure used for cognitive empathy was the perspective taking subscale of the IRI (Davis, 1980). The model constructed to test H1 included Survey (pre, post, follow up), Group (1 or 2), and Survey x Group interaction for fixed effects and participant was included as a random effect factor. Survey was also identified as a repeated effect because each level is repeated for each participant. The Akaike information criterion (AIC) was used to determine the appropriate model. A lower AIC indicates a better model fit; thus, a model with diagonal covariance structure (AIC = 438.34) was chosen when compared to the default variance component structure (AIC = 443.39; Walker et al., 2019). A diagonal covariance structure and a variance component structure both assume no correlation between any elements, but the variance component structure has constant variance, and the diagonal structure has differing variances (Cheng et al., 2008). Other variance structures such as unstructured and compound symmetry were also tested and ruled out.

There was a significant main effect for survey ($\beta = 2.36$, $SE = 0.96$, $t(38.59) = 2.46$, $p = 0.019$) indicating a difference among perspective taking score on the pre, post, and follow-up surveys. Pairwise comparisons revealed that perspective taking scores on the pre and post surveys were significantly different ($MD = -1.79$, $SE = 0.59$, $p = 0.014$). Perspective taking scores on the follow-up survey were not significantly different from pre or post scores. Figure 3

shows that the post scores are significantly higher than pre scores but follow-up scores lowered to a similar level as the pre scores.

Figure 3

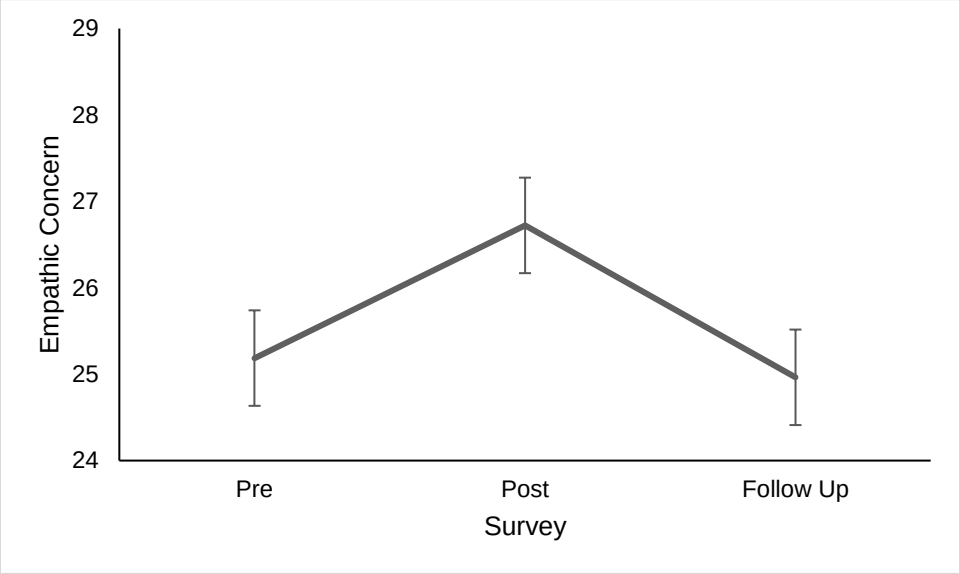
Perspective Taking Mean Scores



A LMM was also constructed to investigate affective empathy across the surveys. The default covariance structure was the best fit for this model based on the AIC and number of parameters and the construction of the model is the same as the model for perspective taking. The measure used for affective empathy was the empathic concern subscale of the IRI (Davis, 1980). The LMM for empathic concern showed significant main effects for survey (β = 2.01, *SE* = 0.92, $t(43.19) = 2.19$, $p = 0.034$) and survey x group interaction (β = 2.87, *SE* = 1.38, $t(43.5) = 2.09$, $p = 0.043$). Pairwise comparisons revealed significant differences between empathic concern scores on the pre and post surveys (*MD* = -1.54, *SE* = 0.56, $p = 0.026$) and marginally significant differences between scores on the post and follow-up surveys (*MD* = 1.76, *SE* = 0.70, $p = 0.050$). Figure 4 shows the same pattern of change as perspective taking.

Figure 4

Empathic Concern Mean Scores



To test the second hypothesis (H2), which predicted that individuals with higher initial empathy (both affective and cognitive) would demonstrate a larger increase in cognitive empathy (not affective empathy) after training than those with lower initial empathy levels, a MANOVA was conducted. The scores from the perspective taking and empathic concern subscales on the pre survey were added together to get a total initial empathy score for each participant and then were labeled as either high or low based on the median score. To assess the impact of the training on post-training empathy levels, participant scores for perspective taking and empathic concern from the post survey were used.

The MANOVA examined initial empathy levels (high and low) as the independent variable and perspective taking and empathic concern as dependent variables. The multivariate result was significant (Wilk's Lambda = 0.47, F (2, 19) = 10.67, $p < 0.001$), indicating a difference in participants' post-training empathy levels based on their initial empathy levels. Univariate tests revealed significant differences on both perspective taking (F(1, 20) = 4.41, p =

0.049, $\eta^2 = 0.18$) and empathic concern scores ($F(1, 20) = 16.30$, $p < 0.001$, $\eta^2 = 0.45$) with post scores higher for individuals with high initial empathy levels. To check whether these differences were sustained after the post survey, a second MANOVA was conducted with the cognitive and affective empathy scores on the follow-up survey. However, no differences were found, possibly due to the very low N ($N = 11$, 4 people with low initial empathy, 7 people with high initial empathy).

Qualitative

There were a number of open-ended questions included on all four surveys. The same or similar questions were repeated on the pre, post, and/or follow-up surveys. A thematic analysis was conducted to analyze the open-ended responses (Braun & Clarke, 2006). The PI was assisted by one graduate and one undergraduate student during the initial coding and theme identification stages. Two new undergraduate students were chosen to assist with reviewing and finalizing the themes. In the initial coding stage, the PI and the assistant coders went through the data separately to familiarize themselves with the data and begin to identify themes. All three then met to discuss their initial thoughts and agree on a list of major themes. The next step was to separately go through the data again to apply the agreed upon themes and identify any missing themes. The final step was to discuss any disagreements and come to a decision on any outliers from the themes. To capture changes in responses to the questions before and after the training, the themes were grouped by question. Patterns of themes were also looked at across Group 1 (spread out sessions) and Group 2 (condensed sessions) but no differences emerged except on Q2 and Q3.

Q1 - My biggest learning from this training:

Q1 was included only on the post and follow-up surveys. On the post survey (total responses on post survey = 25), the majority of responses fell under the theme of learning how to take action against exclusion (44% of post survey responses) with subthemes, learning how to recognize exclusion (4) and realizing the need to take action against exclusion (3). The following quote from a study participant illustrates the theme of learning how to take action against exclusion, “This training has given me some tools to be more aware of situations I might be able to more actively participate in to be more inclusive.” Although this respondent does not seem fully confident yet, they know that they have the tools from the training to intervene and encourage inclusion. This next quote illustrates the subtheme of realizing the need to take action against exclusion, “Ending Racism requires an active role, cannot be a bystander.” This respondent has learned that nothing will change unless people take action to create change. The next most common theme on the post survey was having a deeper understanding of exclusion (36% of post survey responses) followed by a recognition of one’s place on the journey to inclusion (realizing that there is still a lot of work to do toward that goal and that each of us play an important role in it; 28% of post survey responses), as illustrated by this response, “How much I've done wrong. Awareness is painful but promotes growth. I (and the org[anization]) are heading in a positive direction with respect to [equity and inclusion].”

On the follow up survey (total responses on follow up survey = 15), the most common theme was having a deeper understanding of exclusion (40% of follow up responses). The following quotes illustrate this theme: “The accumulation of direct or indirect slights and the frustration that the person doing them does not see them as a slight;” “More awareness about systemic inequity.” These quotes illustrate that respondents learned what it feels like to be the

victim of exclusive behaviors and about the systemic roots of the issue and thus left the training with an overall deeper understanding of exclusion.

Q2 - What are some things that came to mind as you were experiencing this scenario?

Q2 appeared only on the VR survey (total responses to Q2 = 41). The scenario the question refers to is the VR experience/scenario described in the method section. Overall, the most common themes were feelings of frustration (27% of responses) and feeling dismissed (24% of responses). Other common themes overall included recognizing their own experience within the scenario (17% of responses; “I felt emotional reflecting on times that these same scenarios have happened to either myself, my dad, my mom, and my brothers in the corporate world.”) and feeling that exclusion happens a lot more than they realized (17% of responses; “How often things like this could happen and if I'm not looking for them they could slip right by me and I wouldn't even notice them.”).

When comparing the two participant groups (Group 1 = spread out training sessions; Group 2 = condensed trainings sessions), feelings of frustration and feeling dismissed were much more common in Group 1 (frustration = 44%; dismissed = 35%) than in Group 2 (frustration = 6%; dismissed = 11%). There was a noticeable difference among the theme feeling that exclusion happens a lot more than they realized (Group 1 = 4%; Group 2 = 33%). Group 2 also had more responses expressing the need to continue the journey toward inclusion (Group 2 = 17%; Group 1 = 0%), as illustrated by this response: “These types of behaviors happen everywhere, with coworkers, friends, neighbors. Being aware of it is the first step, the harder and more impactful part is how to properly address it in a way that can make meaningful change.”

This respondent recognizes that the behaviors seen in the scenario are indicative of widespread issues and that there is a pressing need to address these issues to take steps toward inclusion.

Q3 - What is something that stood out to you from this experience?

Q3 also appeared only on the VR survey (N = 38). The experience the question refers to is the VR experience/scenario described in the method section. Overall, the most common theme was frustration (37%) with subthemes feeling demotivated (3), feeling unsupported (3), and feeling like exclusion can be frustratingly subtle (4). This last subtheme is illustrated by the quote: “How subtle the discrimination can be but how large of an impact that can have - even if not intentional.” There were slightly more responses under frustration in Group 1 (40%) than Group 2 (33%). The next most common theme for Group 1 was feeling dismissed (15%). The next most common themes for Group 2 were feeling the need to continue the journey toward inclusion (17%) and recognizing their own experience within the scenario (17%).

Discussion

Hypothesis 1 was supported (cognitive empathy was higher after the training compared to before). However, cognitive empathy was predicted to change based on past research suggesting that cognitive empathy is the trainable aspect of empathy whereas affective empathy is the more stable trait (Barlinska et al., 2018; Rueda et al., 2015; van Loon et al., 2018; Zahn-Waxler et al., 1992). The results show a larger change in affective empathy than cognitive empathy after the training compared to before. This could indicate that more research should be done on training and empathy, or it may be an impact that is unique to VR-based training. Past research has shown that VR experiences elicit strong emotional reactions because of the immediacy and realism of the content (Bertrand et al., 2018; Kapp & O’Driscoll, 2010; Schmid Mast et al., 2018). In contrast, most traditional perspective taking exercises ask the participant to imagine

how someone else might feel rather than immediately feeling it themselves. This stronger appeal to participants' emotions could be creating a larger change in affective empathy than traditional perspective taking exercises.

The significant increase among perspective taking and empathic concern scores from pre to post training is encouraging but scores dropped to similar levels as the pre survey just nine weeks after the end of the training on the follow-up survey. The low response rate on the follow-up survey may have affected this outcome but it is consistent with previous research on the impact of DEI trainings. The impact of DEI training usually only lasts a few weeks following the training (Bezrukova et al., 2016; Bezrukova et al., 2012; Paluck & Green, 2009). However, there is a disconnect when looking at this finding alongside the qualitative responses, specifically the responses to Q1 ("My biggest learning from this training:"). The quantitative data shows a decrease in empathy, but the participants report that they still felt they had a deeper understanding of exclusion nine weeks after the training and that they still feel a drive to do something about it. This suggests that while empathy seems to be a key part of making behavioral and attitudinal changes, there may be another factor serving as the long-term motivator for sustaining these changes.

The increase in empathy at the post training stage shown in the quantitative data is reflected in the responses to open-ended Q1 on the post survey. The increase in empathy that the participants reported feeling after the training is likely connected to the realization that action needs to be taken against exclusion and recognizing that each individual has a role to play on the journey to inclusion. The most common theme on the follow up survey was having a deeper understanding of exclusion, which is possibly a result of the increase in empathy following the training. This finding may suggest that the knowledge gained from the training, or the memory

of the feeling of discrimination within the VR experience rather than an increase in empathy per se, is what creates long-term behavioral and attitudinal changes.

The responses to open-ended Q2 (“What are some things that came to mind as you were experiencing this scenario?”) and Q3 (“What is something that stood out to you from this experience?”) demonstrate the strong emotional reactions that participants had to the VR scenario. For either question, the majority of respondents reported feelings of frustration or feeling dismissed as what stood out to them the most from the VR experience. These emotional reactions likely contributed to the increase in empathy following the training. As mentioned in previous research, VR training targets participants’ empathy levels because they have fully experienced something from someone else’s perspective (Bertrand et al., 2018; Herrera et al., 2018; van Loon et al., 2018).

Hypothesis 2 was also supported. There were higher post-training scores for participants with high initial empathy levels on perspective taking scores. However, similar to the results for Hypothesis 1, a larger difference was found on affective empathy: more of the variance in post-training empathic concern scores was explained by initial empathy levels. This finding suggests that what participants are bringing to the training matters. In other words, their own history and personality influences how they receive the training and how impactful it is for them. This has implications for the design and development of diversity training, which is discussed further in the Implications section. Additionally, the unexpected larger differences among empathic concern scores further underscores the need to more closely investigate the impact of VR trainings on affective empathy.

Post-hoc Exploratory Analyses & Discussion

Additional analyses were conducted to examine survey questions that did not directly relate to the primary research questions. The notable findings from those analyses are included in this section. A number of the questions included on the surveys were created by the PI and the founder of Perspectives, who is an expert in organizational training. The post-hoc exploratory analyses investigated the questions related to participant feelings during the VR scenario, inclusive efficacy, intent for action, and similar experiences.

Participant Feelings During the VR Scenario

The VR survey included items that asked about the participants' feelings as Max during the VR scenario. The four items were answered on a scale of 1 (strongly disagree) to 5 (strongly agree). Scores were obtained by totaling the four items, meaning the possible score range was 4-20. The items were: "As Max, I felt confused," "As Max, I felt dismissed," "As Max, I felt that my ability to do my job was being questioned," and "As Max, I started to feel a bit demotivated." The questions were meant to determine if the participants were feeling the way that the scenario was intended to make them feel.

Exploratory analyses were run on these items to look at differences between the two participant groups. Before analyzing these items as one scale, a factor analysis was conducted and found that all of the items loaded onto one factor. This factor explained 63.68% of the variance with an eigenvalue of 2.55. The Kaiser-Meyer-Olkin (KMO) measure was 0.72 and Bartlett's test of sphericity was significant ($\chi 2$ (6) = 59.46, $p < 0.001$). Significant differences were found between Group 1 (spread out training sessions) and Group 2 (condensed training sessions). A reliability analysis on the six items revealed a Cronbach's Alpha of 0.78. An

ANOVA revealed that Group 2 (N = 21, M = 17.62) more strongly agreed with the items on this scale ($F(1, 42) = 4.83$, $p = 0.034$) compared to Group 1(N = 23, M = 15.57).

Only the VR survey revealed any differences between Group 1 and Group 2 in both the quantitative and qualitative data. The differences in the quantitative data can be seen in the group of questions that address feeling specific emotions as Max, the person the participants embody in the VR experience. The qualitative questions (Q2: “What are some things that came to mind as you were experiencing this scenario?” and Q3: “What is something that stood out to you from this experience?”) ask about the participants thoughts during the experience. Both the qualitative and the quantitative questions ask about participants’ thoughts and feelings during the VR experience.

The differences between the two groups are most likely due to the timing of when each group took the VR survey. Group 1 took the VR survey after going through only the exclusion VR scenario and before having any discussion about it. Group 2 took the VR survey after going through both the exclusion and inclusion VR scenarios and having discussions after each experience. It is clear in the responses that Group 2 was able to process the exclusion within the scenario and the participants were starting to come to the final conclusions intended as outcomes of the training: behaviors of exclusion are a common occurrence that is much more frequent for members of certain populations and that everyone can play a part in working to create an inclusive culture. Group 1’s responses are much more emotionally raw than Group 2’s responses and they express much more frustration than Group 2.

The differences in these responses between Groups 1 and 2 gives us a glimpse into what kind of effect the exclusion VR scenario, showing what not to do, might have without the accompanying facilitated discussions or the parallel inclusion scenario, demonstrating what to

do. Participants are left with their raw emotions and no healthy way to process or unpack them. Their responses are much more productive after they have seen what to do in the same situation and have been able to discuss and unpack their emotions with their peers. The group discussions following the exclusion scenario and viewing the inclusion scenario (what to do) seem to be the key to turning the raw emotional responses into a drive to take action to create real change. Therefore, a key conclusion from study 1 is that in order to give people useful and productive tools to create an inclusive culture, they may need to engage in collective conversation that allows them to commit to working toward the same goals and hold each other accountable to those goals. The conversation also gives participants a chance to ask questions and work through their own biases as they come up in discussion. The VR scenario alone is a powerful tool but can only go so far in changing attitudes and behaviors without those added components.

Inclusive Efficacy, Intent for Action, Similar Experiences

The questions relating to inclusive efficacy, intent for action, and similar experiences (to the VR scenario) on the VR survey were developed based on Kraiger and colleagues (1993) application of theories of learning outcomes to training evaluation. The behaviors and situations mentioned in the items refer to the events in the VR scenario that the participants went through before completing this survey. The items were answered on a scale of 1 (strongly disagree) to 5 (strongly agree). The items were initially analyzed as one group of questions using principal component analysis with Oblimin (oblique) rotation. The analysis yielded two factors explaining 73.02% of the variance overall. However, there was one item (“If I saw this behavior in my organization, I would want to address it.”) that accounted for 15.48% of the variance on its own with an eigenvalue of 0.93, almost high enough to be its own factor. This item was only significantly correlated with one other item on the scale and had a very low communality of 0.13.

As a result, this item was removed from the analysis and tested as its own factor, named "intent for action."

Follow-up factor analyses on the five remaining items yielded two factors explaining 85.98% of the variance overall. Table 4 shows the factor loadings and communalities of the items. Factor 1 was named inclusive efficacy and explained 53.84% of the variance with an eigenvalue of 2.69. Factor 2 was named similar experiences and explained 32.14% of the variance with an eigenvalue of 1.61. The Kaiser-Meyer-Olkin (KMO) measure was 0.65 and Bartlett's test of sphericity was significant ($\chi 2$ (10) = 119.95, $p < 0.001$). A reliability analysis on the six items revealed a Cronbach's Alpha of 0.75.

Table 4

Factor Loadings and Communalities for Items Related to Inclusive Efficacy and Similar Experiences

	Factor Loading		
	1	**2**	**Communalities**
This experience helped me to better understand what it feels like to be excluded in the workplace.	**.944**	-.106	.863
I feel this experience will help me better recognize exclusive microbehaviors in my organization.	**.803**	.212	.757
This experience helped me recognize how subtle behaviors that exclude individuals can impact a person's motivation.	**.918**	-.047	.829
I have witnessed or heard about a similar situation in my organization.	.141	**.923**	.924
I have experienced similar situations in my organization.	-.112	**.978**	.926

Note. Factor 1: inclusive efficacy; Factor 2: similar experiences

A MANOVA was conducted with the items relating to inclusive efficacy, intent for action, and similar experiences. The similar experiences questions were removed for being non-

significant and having a very small partial eta squared. The final MANOVA included empathy levels (high and low) as the independent variable and perspective taking and empathic concern scores from the post survey and inclusive efficacy and intent for action scores from the VR survey as dependent variables. The multivariate result was significant (Wilk's Lambda = 0.46, F (4, 17) = 5.07, p = 0.007). Univariate tests revealed significant differences on action (F(1, 20) = 5.38, $p < 0.031$) with individuals with high initial empathy levels indicating that they were significantly more inclined to want to take action against an exclusive behavior that they witnessed. Differences on the inclusive efficacy questions were not significant.

The differences found on the item labeled intent for action indicated that individuals with high initial empathy levels more strongly agreed that they would want to take action if they witnessed exclusive behavior in their workplace than those with low initial empathy levels. This is in line with the findings related to Hypothesis 2. The unique characteristics and histories that individuals bring with them impact how they receive a DEI training, meaning they may get more or less out of the training, based on such individual differences.

Study 2: MEDTAPP

Method

Participants

About 60 healthcare professionals who previously participated in the UCCON MEDTAPP study were sent a survey to complete for the current study. They were contacted via email by the primary investigator of the original MEDTAPP study. We received 27 full responses to the survey. The respondents were 77.8% female, 70.4% White and 18.5% Black or African American. The majority of the respondents were either a Nurse Practitioner (or other APRN; 33.3%) or a Physician (25.9%) and worked in either behavioral health (i.e. community

mental health; 29.6%), a healthcare system affiliated clinic (25.9%), or a hospital setting (22.2%).

Materials and Measures

MEDTAPP Study

UCCON partnered with the Center for Simulations and Virtual Environments Research at the University of Cincinnati (UCSIM) to develop their VR training scenarios. The simulations took about 20 minutes each to go through because the participant was taken through a scene-by-scene debrief of the social determinants of health that were present in the scenario while they were still in the VR headset. One of the simulations featured a man with schizophrenia who has a variety of factors that make him late to his appointment, including a lack of money for transportation. He ends up getting kicked out of his apartment and cannot afford to pay for the prescription medicine ordered by the physician. A second simulation showed an older woman from Puerto Rico who has a variety of chronic health conditions and speaks English with a heavy accent. She struggles to make her appointment because of communication issues and ends up fainting due to her health issues and being taken to the hospital, which leads to another missed appointment. Both simulations were experienced from the first-person perspective. Healthcare professionals participated in one or both simulations and answered questions before and after the training on things such as their level of empathy and how comfortable they felt interviewing clients of different cultural backgrounds.

Current Study

For the current study, a short survey was developed and sent out via email using Qualtrics. Questions on the survey addressed how the MEDTAPP training impacted the healthcare professionals' interactions with patients, how the MEDTAPP training compared to

other more traditional diversity and inclusion trainings that do not use VR, and demographic questions to capture if there are differences in training impact along the lines of job type or amount of time spent with patients. The full survey is included in Appendix B.

Design and Procedure

The primary investigator of the original MEDTAPP study sent out an email in February 2021 to the healthcare professionals who previously participated in the MEDTAPP training. The email reminded them of the study they participated in, explained the current study, and asked them to follow a link to the survey for the current study. All responses were received by March 2021.

Results

Quantitative

To test the quantitative questions on the survey, a series of MANOVAs were conducted. The demographic questions (job title, time spent with patients, work setting, gender, and race) were tested as independent variables (IV) in separate MANOVAs. The six Likert-style questions and the dichotomous questions ("Have you changed the way you approach your patients since the VRS?" and "Had you participated in other diversity/inclusion training(s) about social determinants of health before this one?") were included as dependent variables (DV) in each of the MANOVAs.

None of the MANOVAs revealed any significant differences. However, follow-up ANOVAs were conducted on items with lower alpha levels (closer to $p = 0.05$) on the univariate tests to look more closely at any trending differences. An ANOVA with time spent with patients as the IV and the item "As a result of the virtual reality simulation (VRS), my relationships with my patients have improved" as the DV showed potential differences between participants who

spent two hours a day with patients and those who spent five or more hours a day with patients ($F(5, 16) = 2.57$, $p = 0.069$). An ANOVA with Gender as the IV and the item "The VRS made me aware of social determinants of health that I did not previously know of" as the DV was significant ($F(1, 25) = 5.92$, $p = 0.022$). Females ($N = 21$, $M = 4.00$) more strongly agreed with this question than males ($N = 6$, $M = 2.83$). There were not enough participants within each racial group to find meaningful differences. Race was also tested as a binary variable (non-white and white) and no meaningful differences were found.

A factor analysis was conducted on the six Likert-style items on the survey to determine if they could be tested as one factor. The six items were analyzed using principal component analysis with Oblimin (oblique) rotation. The analysis yielded one factor explaining 72.79% of the variance with an eigenvalue of 4.37. The Kaiser-Meyer-Olkin (KMO) measure was 0.86 and Bartlett's test of sphericity was significant ($\chi 2$ (15) = 113.34, $p < 0.001$). The factor was named "overall impact of the training."

ANOVAs with each of the demographic variables (job title, time spent with patients, work setting, gender, race, and race binary) as the IVs and overall impact as the DV were conducted. No significant differences were found among any variables.

Respondents were asked to input the approximate date that they participated in the MEDTAPP training. This variable was transposed into time in months since the training. The average length of time from when respondents took the training to when they took the survey was about 18 months. The maximum amount of time was 26 months and the minimum amount of time was 12 months. To be able to identify patterns better, the time elapsed since training was grouped into five response categories (12-14 months, 15-17 months, 18-20 months, 21-23

months, and 24-26 months). An ANOVA was conducted with time since training as the IV and overall impact as the DV. No significant differences were found.

Qualitative

The main goal of Study 2 was to gain insight into the impact of the training on participants. A thematic analysis was conducted to analyze the open-ended responses (Braun & Clarke, 2006). The PI was assisted by one graduate and one undergraduate student in the analysis process. The same steps were followed to analyze these qualitative responses as were followed in Study 1. The themes were grouped by question to analyze each one closely. In the survey questions, “VRS” refers to the “virtual reality simulation project,” also referred to as the MEDTAPP training in this paper.

Q1 - If yes, how have you changed the way you approach your patients?

Q1 was a follow-up to the yes or no question: “Have you changed the way you approach your patients since the VRS?” The majority of respondents to Q1 (47% of the total responses; total responses = 19) described having increased consideration of their patient’s circumstances and taking this into account when determining the best course of care. This theme is illustrated by the following quotes from study participants: “Try to consider beyond the surface, those complex issues that might impact patient compliance;” “considering all aspects of a patient's circumstances to better determine my clinical approach." The respondents demonstrate that they spend more time thinking about their patient’s circumstances so that they can tailor each patient’s plan of care.

Q2 - If you are able, please tell us a story about a patient interaction you think was more positive as a result of the VRS.

The majority of the responses to Q2 (63% of the total responses; total responses = 8) mentioned being able to better identify and overcome barriers to healthcare for their patients. This theme is illustrated by the following quotes: (Respondent A) "An elderly lady presented with limited healthcare literacy, limited access to meds, poverty, poor / non-existent home support, and food insecurity. She was at a tipping point when she presented to the [emergency department]. I was able to identify the pertinent issues and coordinate with case management and social work to get her the appropriate support and care. I believe she would have suffered a very poor outcome had these factors not been acutely recognized and addressed;" (Respondent B) "Overall, changing approach when miss appointment. Inquire if they are ok or had an issue before talking about plan to improve attendance."

Q3 - How would this interaction have gone before the VRS and how did it go after the VRS?

Q3 was a follow up question to Q2. Respondents to Q3 (total responses = 8) reported either that they would have behaved differently (50%) or that they would have made incorrect assumptions about the patient without talking about it (38%). The quotes included here for Q3 correspond to the quotes included in Q2. The following quote illustrates the theme that they would have behaved differently before the training: (Respondent A) "I may have simply discharged her with an outpatient referral and the inappropriately optimistic hopes that follow up would occur in a timely fashion and a crisis would not occur in the meantime." The next quote illustrates the theme of making incorrect assumptions before the training: (Respondent B) "Maybe making assumption that they should have called, rather than inquiring." Looking at the

responses to both Q2 and Q3 gives a clear picture of how the healthcare professionals changed their approach to patients before and after the training.

Q4 - If yes, how did the experience and impact of the VRS compare to the prior trainings you participated in?

Q4 was a follow up to the question: "Had you participated in other diversity/inclusion training(s) about social determinants of health before this one?" Responses to Q4 (total responses = 14) included themes like the training was more engaging and immersive (43%), as illustrated by this quote: "It was much cooler and much more engaging and exciting to use." A few participants reported having emotional responses to the experience and feeling that it was more realistic, as illustrated by the following quotes: "Far better, I felt it," and, "I think this one was better because it gave a better sense of the actual day to day reality of our patients rather than just someone lecturing about what globally impacts patient outcomes." Some reported being excited to use the VR technology. One respondent suggested that the training would have been more impactful if it was more than just one day.

Q5 - Overall, what impact did the VRS have on you as a healthcare professional?

In the responses to Q5 (total responses = 23), the most reported change was an increased consideration for patients' circumstances, which changed how they behave towards patients (26%), as exemplified by this quote: "As a healthcare professional, the VRS helped me to better understanding the clients I serve in relation to events happening outside of the clinic setting which may [have] greatly impacted the services I provide to them." Participants also reported having more awareness of social determinants of health (22%) and having more empathy or compassion for patients with barriers to healthcare (22%), as illustrated by the following quote: "I've gained significant insight into SDOH and a gut-level feel for their impact. Previously, I

simply had the knowledge without the more profound empathy that this program helped me develop. I believe that I'm a better doctor for my patients now."

Discussion

Although the small number of respondents limited possible results when examining the quantitative data, there are some findings worth noting. Respondents who reported spending more of their workday with patients more strongly agreed that the relationships with their patients had improved because of the training. This finding may suggest that participants who had more opportunities to apply what they had learned in the training may have experienced more of an impact from the training compared to those with fewer opportunities to apply their learning.

When looking at overall impact and time since training, there were no significant differences between those who participated in the training 12 months ago and those who participated over two years ago. This indicates that there was no downward trend in impact as time passed. The shortest amount of time since completing the training was 12 months for any of the respondents and yet many of them still attributed changes in their patient care to this one-day training. Participants also did not seem to have any trouble recalling the content of the training from over a year ago. This speaks to the lasting impression of a virtual reality experience. Participants reported having embodied emotional reactions to the experience (e.g. "I felt it"), which means that the experience will stay with them longer than if they were just told about patients' experiences with social determinants of health.

General Discussion

In this section, the results of Studies 1 and 2 are taken together to identify trends and future research questions. The first theme identified across the two studies is the lasting impact

VR training can have on participants. Although the quantitative measures of empathy in Study 1 fell to pre-training levels nine weeks after the training, the qualitative responses demonstrated that there was a lasting impact from the training. The participants reported continuing to recognize exclusive acts and feel motivated to act toward eradicating exclusion because they understood the negative impact of exclusion. A lasting impact can also be seen in the qualitative responses from Study 2. This effect is likely connected to the strong emotional responses the participants had to the VR experiences. Participants from both studies reported remembering what it felt like to be in the situation within the VR experience. That feeling serves as a reminder of what they learned in the training when they encounter similar issues in real life. The memory of this feeling seems to be a motivating factor that continues past the point when empathy levels have returned to baseline.

A second theme from the two studies is how to create change at the organizational level. The training in Study 2 had no discussion component after the VR experience, only a short debrief of concepts built into the simulation, such as highlighting where housing stability or isolation is presented within the scenario. The results from Study 1 show the impact that having discussions to unpack your reactions can have on learning and motivation but the findings from Study 2 demonstrate that a fully virtual training with no in-person facilitator, can still be impactful. Organizations do not necessarily have to invest in a long training with a facilitator to create change within their workplace. However, the changes reported in Study 2 are changes at the individual level. The findings from Study 1 suggest that to create change at the organizational level, it is likely necessary to have some component of the training involve peer discussions to get everyone working together towards the same goal of improving the company as a whole.

A third theme identified across the findings of Study 1 and Study 2 is the possible advantages and disadvantages of using VR technology within workplace training. Many of the participants from both studies showed great enthusiasm for the training because of the VR experience which may be a reason to implement VR-based training over more traditional training methods. Because VR is still very new to the workplace and employees get excited to try out this new technology, it is quite possible they are more engaged in the training than they might be otherwise, especially if the topics are familiar to them from past DEI training programs. On the other hand, this could also be a barrier to employees participating if they are intimidated by the technology.

Implications

The findings from the current studies inform important aspects of implementing a VR-based DEI training or DEI training in general that organizations and their leaders should keep in mind. First, organizational leaders should be aware that the extent of change they can expect from a training depends on the length and depth of the training they choose. To create individual-level behavior or attitudinal change, organizations can choose an entirely virtual, one-day training. However, to create organizational-level change, they will likely need to invest in a longer training that includes facilitated peer discussions that have everyone work toward a collective goal.

Second, when choosing a VR-based training, organizational leaders need to remember that the technology could be something that makes employees more likely to sign up for the training or something that makes employees less likely to sign up if they are uncomfortable with new technology like VR. Leaders should know who they are providing the training for and choose an appropriate platform for their audience.

Third, it may be helpful when implementing DEI training programs for leaders to take into account individual differences in the reception and experience of a training. A facilitator should understand that some people will be more open to and engaged in the training material than others depending on their personality or other factors, but this does not mean that those who are less open are not still benefitting from the training. The impact may be less apparent in these individuals but there is still an impact.

Fourth, the way to create motivation for action is to give participants an outlet through which to process their emotions and show them what to do in order to give them the tools to create productive change. This outlet could be in the form of facilitated peer discussions or it could be a guided debrief of the concepts illustrated within the training. Without these components, the trainings may do more harm than good by leaving the individual to come to their own conclusions about what they saw in the training scenario and decide on their own course of action.

Fifth, leaders need to appreciate the importance of creating a practice of open dialogue and discussion about difficult topics, such as exclusion, especially exclusion happening within one's own workplace. A practice of open dialogue goes a long way toward creating lasting culture change where employees are comfortable voicing any issues and feel that their voice is being heard. Creating a culture where employees are comfortable speaking out and feel well-received by their peers and leaders leads to increased productivity and commitment from employees (Shore et al., 2018).

Finally, organizational leaders and training facilitators should keep in mind that employees need opportunities to apply what they have learned in a training to create long-term learning retention. Many of the healthcare workers who participated in Study 2 worked daily

with patients with similar circumstances to the ones portrayed in the VR scenario. As a result, Study 2 participants were able to apply what they learned in the training to their work daily. In contrast, the engineers who participated in Study 1 only got the opportunity to apply their learnings occasionally, if at all. Research has shown that being able to apply new skills after learning them is crucial to long-term learning retention (Mattingly & Kraiger, 2019; Kapp & O'Driscoll, 2010).

Limitations & Future Directions

A limitation of the current research is that there was a lack of variance in the demographics in both studies, which limited the conclusions that could be drawn from the data. A key difference between the two studies that should be noted is the distribution of gender within the participants of each study. Study 1 had a majority of male participants, while Study 2 had a majority of female participants. Research has shown that women inherently have higher empathy levels than men, making it likely that they will respond differently to a training that targets empathy (Chen et al., 2014; Christov-Moore et al., 2014; Eisenberg & Lennon, 1983; Rueckert & Naybar, 2008). The differences in demographics between the two studies may have contributed to changes in the findings that would not have been present with more balanced participant groups.

The data collected for both studies was all self-report, which limits the scope of the conclusions. Future research with more diverse populations and data from additional sources, perhaps from coworkers or leaders of the training participants, would strengthen our insight into the impact of this type of training.

The was a much lower response rate on the follow-up surveys in Study 1 compared to the other surveys. This may have been due to the timing of the surveys, which went out during the

summer when many people were taking vacation. For future studies, this could be improved by being mindful of the timing of these surveys or having a leader within the organization send out the surveys, rather than the training facilitator who is outside the company, to encourage more responses to the surveys. The low response rate on the follow up survey may have contributed to the pattern seen in both of the empathy measures. Both affective and cognitive empathy scores increased significantly from pre survey to post survey and decreased almost the same amount nine weeks later on the follow-up survey. Since this pattern was seen across multiple measures (cognitive empathy and affective empathy), it may point to a flaw in the design of the study rather than specific variable-level changes. However, the measures used for cognitive and affective empathy were moderately correlated, which may have limited the differences discovered in the effect of the training on these variables.

In Study 1, there were multiple parties involved in the training experience (the organization participating in the training, the company providing the training, and the PI of the current study) who all had their own goals for what they wanted to get out of the training. Because of all of these interests, the survey questions had to be adjusted to give space multiple goals. The survey questions included in Appendix A do not include extra items that were added by the organization participating in the training. These extra items were specific to previous internal diversity initiatives the organization had been involved in and thus were not evaluated as part of the current research. However, these questions did limit the number of questions that could be included in the surveys to avoid survey fatigue. Having to accommodate for multiple parties' goals is a natural side effect of conducting research within functioning organizations.

A possible limitation for Study 1 is that peer discussions may have influenced individual's thoughts or feelings, leading them to all come to the same conclusions. Perhaps if

individual participants were left to come to their own conclusions, we would have seen more variety in responses to the training. However, it would be risky to leave each individual to come to their own conclusion and decide what they are going to do moving forward. For example, participants could conclude that it is hopeless and there is no point trying to do anything or they may decide that since the intention behind the subtle acts of exclusion was not bad, the negative impact should not count or is not something to be worried about. These conclusions serve to create more harm than good because they permit people to continue to do nothing. This inaction allows the problem to continue unchecked by society and eventually grow to create irreparable issues. By having facilitated discussions with their peers, participants are able to be led toward productive conclusions and create their own action plans along with their peers to hold each other accountable. Further investigation on the effects of peer discussions would help to provide insight into how such large group discussions influence training impact.

The finding from Study 1 that affective empathy had a more dramatic change than cognitive empathy after the training is something that future research should investigate. Further investigation would help to determine whether this finding is something specific to a VR training format or perhaps researchers do not know as much about these two aspects of empathy as previously thought. A direct comparison of a VR-based and traditional DEI training program would help determine if there are different effects on empathy based on training format. A direct comparison study would also demonstrate whether attitudinal and behavioral changes on other measures are more pronounced following one type of training compared to the other.

Another concept that future research could investigate is whether there is a certain amount of time following training that training impact begins to plateau. The results from Study 2 did not show a change in impact over time starting with one year after training. Is there a point

prior to a year after training when impact begins to plateau and are those measures of impact higher than before the participants went through the training? This question should be examined in future research on the topic.

Finally, in future research, quantitative measures in addition to empathy should be gathered to capture the impact of the training. It is clear from the qualitative responses from both studies that there is lasting impact from the VR trainings, but it is not clearly shown in the quantitative measures. Because organizations are more inclined to be convinced by numbers, it is important to find measures that reflect the true impact of a VR-based DEI training. In other words, measures that discern the true motivating factor behind any attitudinal or behavioral changes reported in the qualitative questions. The questions that result from these findings are important. Clearly there is meaningful impact occurring that we as researchers are not adequately capturing. Although we are not yet able to put a name to it, something beyond empathy is motivating individuals to continue to change their mindset and their behavior long after the training. This could be continued application or reminders of learned knowledge, the memory of how they felt during the VR experience, or a combination of these factors. Whatever this motivating factor is, in order to really move diversity training forward, we as researchers need to devote considerable attention to deciphering it.

Conclusion

While there is still much that we need to learn about using VR for DEI training, the lasting response from participants points to many benefits for this training platform. Many organizations are taking steps to create more inclusive workplaces, but some may be approaching it in a way that creates more problems. Additional research on the impact of DEI trainings can help to inform organizations in choosing a beneficial type of training as well as discussions and

other follow-up activities. A carefully chosen, well-implemented training can be the difference between employees who feel accepted and safe in their workplace and employees who feel othered or tokenized.

References

Anand, R., & Winters, M. F. (2008). A retrospective view of corporate diversity training from 1964 to the present. *Academy of Management Learning & Education*, *7*(3), 356-372.

Arthur Jr, W., Bennett Jr, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, 88(2), 234.

Asare, J. G. (2019, March 31). 5 reasons why diversity programs fail. Forbes. https://www.forbes.com/sites/janicegassam/2019/03/31/5-reasons-why-diversity-programs-fail/?sh=5785bc2a637d

Barlińska, J., Szuster, A., & Winiewski, M. (2018). Cyberbullying among adolescent bystanders: role of affective versus cognitive empathy in increasing prosocial cyberbystander behavior. *Frontiers in Psychology*, 9, 799.

Batson, C. D., Sympson, S. C., Hindman, J. L., Decruz, P., Todd, R. M., Weeks, J. L., ... & Burns, C. T. (1996). "I've been there, too": Effect on empathy of prior experience with a need. *Personality and Social Psychology Bulletin*, 22(5), 474-482.

Bendick Jr, M., Egan, M. L., & Lofhjelm, S. M. (2001). Workforce diversity training: From anti-discrimination compliance to organizational development. *Human Resource Planning*, 24(2).

Bertrand, P., Guegan, J., Robieux, L., McCall, C. A., & Zenasni, F. (2018). Learning empathy through virtual reality: multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Frontiers in Robotics and AI*, 5, 26.

Bethlehem, R. A., Allison, C., van Andel, E. M., Coles, A. I., Neil, K., & Baron-Cohen, S. (2017). Does empathy predict altruism in the wild?. *Social Neuroscience*, 12(6), 743-750.

Bezrukova, K., Jehn, K. A., & Spell, C. S. (2012). Reviewing diversity training: Where we have been and where we should go. *Academy of Management Learning & Education*, 11(2), 207-227.

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*, 142(11), 1227.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Campbell, J. C., Hays, M. J., Core, M., Birch, M., Bosack, M., & Clark, R. E. (2011). Interpersonal and leadership skills: using virtual humans to teach new officers. In Proc. of Interservice/Industry Training, Simulation, and Education Conference, Paper (Vol. 11358).

Cane, J. E., Ferguson, H. J., & Apperly, I. A. (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 591-610.

Catalyst. (2020, June 24). Why diversity and inclusion matter: Quick take. Catalyst. https://www.catalyst.org/research/why-diversity-and-inclusion-matter/

Chang, E. H., Milkman, K. L., Zarrow, L. J., Brabaw, K., Gromet, D. M., Rebele, R., Massey, C., Duckworth, A. L., & Grant, A. (2019, July 9). Does diversity training work the way it's supposed to?. Harvard Business Review. https://hbr.org/2019/07/does-diversity-training-work-the-way-its-supposed-to

Chang, J. W., Sy, T., & Choi, J. N. (2012). Team emotional intelligence and performance: Interactive dynamics between leaders and members. *Small Group Research, 43*(1), 75-104.

Chen, W., Lu, J., Liu, L., & Lin, W. (2014). Gender differences of empathy. *Advances in Psychological Science, 22(9),* 1423.

Cheng, J., Edwards, L. J., Maldonado-Molina, M. M., Komro, K. A., & Muller, K. E. (2010). Real longitudinal data analysis for real people: building a good enough mixed model. *Statistics in medicine, 29*(4), 504-520.

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & biobehavioral reviews*, *46*, 604-627.

Combs, G. M. (2002). Meeting the leadership challenge of a diverse and pluralistic workplace: Implications of self-efficacy for diversity training. *Journal of Leadership Studies*, 8(4), 1-16.

Combs, G. M., & Luthans, F. (2007). Diversity training: Analysis of the impact of self-efficacy. *Human Resource Development Quarterly*, 18(1), 91-120.

Costello, A. and Osborne, J. (2005). Exploratory Factor Analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(7), 1-9.

Cox, C. B., House, A., Lopez, A., & Pool, G. J. (2017). The Need for Conceptual Models of Technology in Training and Development: How Immersive Does Training Need to Be?. *Industrial and Organizational Psychology*, 10(4), 696-701.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113.

De la Peña, N., Weil, P., Llobera, J., Giannopoulos, E., Pomés, A., Spanlang, B., ... & Slater, M. (2010). Immersive journalism: immersive virtual reality for the first-person experience of news. Presence: Teleoperators and virtual environments, 19(4), 291-301.

De Waal, F. B. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, 59, 279-300.

Deaton, J. E., Barba, C., Santarelli, T., Rosenzweig, L., Souders, V., McCollum, C., Seip, J., Knerr, B. W., & Singer, M. J. (2005). Virtual environment cultural training for operational readiness (VECTOR). *Virtual Reality*, 8(3), 156-167.

Decety, J., & Yoder, K. J. (2016). Empathy and motivation for justice: Cognitive empathy and concern, but not emotional empathy, predict sensitivity to injustice for others. *Social Neuroscience*, 11(1), 1-14.

Dobbin, F., & Kalev, A. (2016, July-August). Why diversity programs fail: And what works better. Harvard Business Review. https://hbr.org/2016/07/why-diversity-programs-fail

Doe, R., Ndinguri, E., & Phipps, S. T. (2015). Emotional intelligence: The link to success and failure of leadership. *Academy of Educational Leadership Journal*, 19(3), 105.

Duguid, M. M., & Thomas-Hunt, M. C. (2015). Condoning stereotyping? How awareness of stereotyping prevalence impacts expression of stereotypes. *Journal of Applied Psychology*, 100(2), 343–359.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, *94*(1), 100–131.

Fitzmaurice, G. M., & Ravichandran, C. (2008). A primer in longitudinal data analysis. *Circulation*, 118(19), 2005-2010.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, 78(4), 708.

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., Smith, C. D., & Satava, R. M. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery*, 241(2), 364.

Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., & Tecchia, F. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6), 778-798.

Groves, K. S., McEnrue, M. P., & Shen, W. (2008). Developing and measuring the emotional intelligence of leaders. *Journal of Management Development*. 27(2), 225-250.

Gurchiek, K. (2020, December 17). Businesses respond to demand to put muscle into DE&I efforts. Society for Human Resource Management. https://www.shrm.org/ResourcesAndTools/hr-topics/behavioral-competencies/global-and-cultural-effectiveness/Pages/Businesses-Respond-to-Demand-to-Put-Muscle-into-DEI-Efforts.aspx

Hauenstein, N. M., Findlay, R. A., & McDonald, D. P. (2010). Using situational judgment tests to assess training effectiveness: Lessons learned evaluating military equal opportunity advisor trainees. *Military Psychology*, 22(3), 262-281.

Hays, M. J., Campbell, J. C., Trimmer, M. A., Poore, J. C., Webb, A. K., & King, T. K. (2012). Can role-play with virtual humans teach interpersonal skills?. University of Southern California Los Angeles Inst For Creative Technologies.

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PloS one*, 13(10), e0204494.

Holladay, C. L., Knight, J. L., Paige, D. L., & Quiñones, M. A. (2003). The influence of framing on attitudes toward diversity training. *Human Resource Development Quarterly*, 14(3), 245-263.

Hsu, E. B., Li, Y., Bayram, J. D., Levinson, D., Yang, S., & Monahan, C. (2013). State of virtual reality based disaster preparedness and response training. *PLoS currents*, 5.

Kapp, K. M., & O'Driscoll, T. (2009). *Learning in 3D: Adding a new dimension to enterprise learning and collaboration*. John Wiley & Sons.

Karagiorgi, Y., & Symeou, L. (2005). Translating constructivism into instructional design: Potential and limitations. *Journal of Educational Technology & Society*, 8(1), 17-27.

Kelly, E., & Dobbin, F. (1998). How affirmative action became diversity management: Employer response to antidiscrimination law, 1961 to 1996. *American Behavioral Scientist*, *41*(7), 960-984.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78(2), 311.

Krebs, D. (1975). Empathy and altruism. *Journal of Personality and Social Psychology*, 32(6), 1134-1146.

Krueger, C., & Tian, L. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing*, 6(2), 151–157. https://doi.org/10.1177/1099800404267682

Kulik, C. T., Pepper, M. B., Roberson, L., & Parker, S. K. (2007). The rich get richer: Predicting participation in voluntary diversity training. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 28(6), 753-769.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19(1), 42-58.

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, 14(1), 27-50.

Lindsey, A., King, E., Hebl, M., & Levine, N. (2015). The impact of method, motivation, and empathy on diversity training effectiveness. *Journal of Business and Psychology*, 30(3), 605-617.

Lindsey, A., King, E., Membere, A., & Cheung, H. K. (2017, July 28). Two types of diversity training that really work. Harvard Business Review. https://hbr.org/2017/07/two-types-of-diversity-training-that-really-work

Luu, T. T., Rowley, C., & Vo, T. T. (2019). Addressing employee diversity to foster their work engagement. *Journal of Business Research*, *95*, 303-315.

Madera, J. M., Neal, J. A., & Dawson, M. (2011). A strategy for diversity training: Focusing on empathy in the workplace. *Journal of Hospitality & Tourism Research*, 35(4), 469-487.

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, 66(5), 1141-1164.

Mattingly, V., & Kraiger, K. (2019). Can emotional intelligence be trained? A meta-analytical investigation. *Human Resource Management Review*, 29(2), 140-155.

Meadows, A., Higgs, S., Burke, S. E., Dovidio, J. F., van Ryn, M., & Phelan, S. M. (2017). Social dominance orientation, dispositional empathy, and need for cognitive closure moderate the impact of empathy-skills training, but not patient contact, on medical students' negative attitudes toward higher-weight patients. *Frontiers in Psychology*, 8, 504.

Misangyi, V. F., LePine, J. A., Algina, J., & Goeddeke Jr, F. (2006). The adequacy of repeated-measures regression for multilevel research: Comparisons with repeated-measures ANOVA, multivariate repeated-measures ANOVA, and multilevel modeling across various multilevel research designs. *Organizational Research Methods*, *9*(1), 5-28.

Noon, M. (2018). Pointless diversity training: Unconscious bias, new racism and agency. *Work, Employment and Society*, 32(1), 198-209.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5), 481-494.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, 60, 339-367.

Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43(1-2), 103-121.

Rabl, T., del Carmen Triana, M., Byun, S. Y., & Bosch, L. (2020). Diversity management efforts as an ethical responsibility: How employees' perceptions of an organizational integration and learning approach to diversity affect employee behavior. *Journal of Business Ethics*, *161*(3), 531-550.

Ragins, B. R., & Ehrhardt, K. (2020). Gaining perspective: The impact of close cross-race friendships on diversity training and education. *Journal of Applied Psychology*. Advance online publication. https://doi.org/10.1037/apl0000807

Rossnagel, C. (2000). Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, 30(3), 429-445.

Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and cognition, 67*(2), 162-167.

Rueda, P., Fernández-Berrocal, P., & Baron-Cohen, S. (2015). Dissociation between cognitive and affective empathy in youth with Asperger Syndrome. *European Journal of Developmental Psychology*, 12(1), 85-98.

Salas, E., Rosen, M. A., Burke, C. S., & Goodwin, G. F. (2008). The wisdom of collectives in organizations: An update of the teamwork competencies. In *Team effectiveness in complex organizations* (pp. 73-114). Routledge.

Schmid Mast, M., Kleinlogel, E. P., Tur, B., & Bachmann, M. (2018). The future of interpersonal skills development: Immersive virtual reality training with virtual humans. *Human Resource Development Quarterly*, 29(2), 125-141.

Shih, M., Wang, E., Trahan Bucher, A., & Stotzer, R. (2009). Perspective taking: Reducing prejudice towards general outgroups and specific individuals. *Group Processes & Intergroup Relations*, 12(5), 565-577.

Shore, L. M., Cleveland, J. N., & Sanchez, D. (2018). Inclusive workplaces: A review and model. *Human Resource Management Review*, 28(2), 176-189.

Smith, T. W. (2008). Loving and caring in the United States: Trends and correlates of empathy, altruism, and related constructs. In B. Fehr, S. Sprecher, & L. G. Underwood (Eds.), *The science of compassionate love: Theory, research, and applications* (p. 81–119). Wiley Blackwell.

Sorkin, A. R., Karaian, J., de la Merced, M. J., Hirsch, L., & Livni, E. (2020, December 1). Nasdaq pushes for diversity in the boardroom. The New York Times. https://www.nytimes.com/2020/12/01/business/dealbook/nasdaq-diversity-boards.html

Sturdivant, M., Yibass, S., Abraham, E., & Hauenstein, N. M. (2017). Using Situational Judgment Tests To Study Subtle Discrimination. *Industrial and Organizational Psychology*, 10(1), 94.

U.S. Department of Health and Human Services. (n.d.). *Social Determinants of Health*. Social Determinants of Health - Healthy People 2030. Retrieved March 9, 2022, from https://health.gov/healthypeople/objectives-and-data/social-determinants-health

van Loon, A., Bailenson, J., Zaki, J., Bostick, J., & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PloS one*, 13(8).

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: The mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology*, 33(4), 455-472.

Perspectives. (2020, August 07). Retrieved January 14, 2021, from http://vrperspectives.com/

Walker, E. A., Redfern, A., & Oleson, J. J. (2019). Linear mixed-model analysis to examine longitudinal trajectories in vocabulary depth and breadth in children who are hard of hearing. *Journal of Speech, Language, and Hearing Research*, *62*(3), 525-542.

Westerman, A., Benk, R., & Greene, D. (2020, December 30). In 2020, protests spread across the globe with a similar message: Black lives matter. National Public Radio. https://www.npr.org/2020/12/30/950053607/in-2020-protests-spread-across-the-globe-with-a-similar-message-black-lives-matt

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and psychological measurement*, *76*(6), 913–934. https://doi.org/10.1177/0013164413495237

Wyld D.C. (2010) A Virtual Explosion or SNAFU is Always Better Than a Real One: Exploring the Use of Virtual Worlds for Simulation and Training…and Developing the Leaders of

Tomorrow. In Iskander M., Kapila V., Karim M. (eds) *Technological Developments in Education and Automation*. Springer, Dordrecht.

Zahn-Waxler, C., Robinson, J. L., & Emde, R. N. (1992). The development of empathy in twins. *Developmental Psychology*, 28(6), 1038–1047.

Zalis, S. (2019, November). Inclusive ads are affecting consumer behavior, according to new research. Think with Google. https://www.thinkwithgoogle.com/future-of-marketing/management-and-culture/diversity-and-inclusion/thought-leadership-marketing-diversity-inclusion/

Appendix A

Full Surveys for Study 1: Perspectives

Pre-Training Survey (taken one-week before the training)

The following are answered on a scale of 1 (Strongly disagree) to 5 (Strongly agree):

1. I think I have high situational awareness when it comes to team dynamics.
2. I think I have high social and emotional intelligence when it comes to team dynamics.
3. I understand the value of increasing my social emotional competency.
4. I feel like I am pretty aware of the cultural nuances that take place in conversations on my team.
5. I feel comfortable to point out moments of exclusion I hear in conversations on my team.
6. My team culture supports and encourages difficult conversations.

Empathy

7. Please choose how well each of the following statements describes you in general on a scale from 1 (does not describe me at all) to 5 (describes me very well). [These items are taken from the Interpersonal Reactivity Index (IRI; Davis, 1980). They include the perspective taking and empathic concern subscales.]

    a. I often have tender, concerned feelings for people less fortunate than me.
    b. I try to look at everybody's side of a disagreement before I make a decision.
    c. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (-)
    d. I sometimes try to understand my friends better by imagining how things look from their perspective.
    e. I would describe myself as a pretty soft-hearted person.
    f. If I'm sure I'm right about something, I don't waste time listening to other people's arguments. (-)
    g. When I'm upset at someone, I usually try to "put myself in their shoes" for a while.
    h. When I see someone being taken advantage of, I feel kind of protective toward them.
    i. I am often quite touched by things that I see happen.
    j. Before criticizing somebody, I try to imagine how I would feel if I were in their place.
    k. Sometimes I don't feel sorry for other people when they are having problems. (-)
    l. I believe there are two sides to every question and try to look at them both.
    m. Other people's misfortunes do not usually disturb me a great deal. (-)
    n. I sometimes find it difficult to see things from the "other person's" point of view. (-)

8. In the past month, how often have you noticed subtle acts of exclusion in the workplace?
    - Often
    - Sometimes
    - Rarely
    - Never

9. When you witnessed these acts of exclusion, how did you respond?
    - I did nothing
    - Confronted the person who did the act of exclusion
    - Tried to smooth things over in the moment
    - Offered support to the person who may have been excluded, privately
    - Called out the behavior in a way that didn't single anyone out

10. Do you feel that your response was appropriate?
11. Why do you feel that your response was (or was not) appropriate?
12. If you did not respond, do you wish that you had? How do you wish you had responded?

13. In the past month, have you had any conversations about gender or racial bias?
    - Often
    - Sometimes
    - Rarely
    - Never

14. Within the last month, what specific actions did you take to make people around you feel more included? Please list them here.

Demographics: The following question group relates to demographic information. **Optional.**

1. How would you describe your disability / ability status? (mark all that apply)
    a. A sensory impairment (eg. hearing / vision)
    b. A learning disability (eg. dyslexia, ADHD)
    c. A long-term medical illness (eg. epilepsy, cystic fibrosis)
    d. A mobility impairment
    e. A mental health disorder
    f. A temporary impairment due to illness or injury (broken ankle, surgery)
    g. I do not identify with a disability or impairment
    h. I prefer not to answer

2. What is your gender? (Mark all that apply)
    a. Male
    b. Female
    c. Nonbinary / Third Gender
    d. Transgender
    e. Cisgender
    f. Agender
    g. Genderqueer
    h. I prefer not to answer
    i. Other:_____

3. Which categories describe you? (Mark all that apply)

   a. American Indian or Alaskan Native
   b. Hispanic, Latino or Spanish origin
   c. Asian
   d. Middle Eastern or North African
   e. Black or African American
   f. Native Hawaiian or other Pacific Islander
   g. White
   h. I prefer not to answer
   i. Other:_____
4. What is your age range?
   a. 18-24 years old
   b. 25-34 years old
   c. 35-44 years old
   d. 45-54 years old
   e. 55-64 years old
   f. 65+ years old
   g. Prefer not to answer

VR Survey (taken after VR experiences)

1. What are some things that came to mind as you were experiencing this scenario?
2. In one word, what do you feel right now:

3. Please choose how much you agree with each statement on a scale of 1 (strongly disagree) to 5 (strongly agree).

   a. As Max, I felt confused.
   b. As Max, I felt dismissed.
   c. As Max, I felt that my ability to do my job was being questioned.
   d. As Max, I started to feel a bit demotivated.
   e. This experience helped me to better understand what it feels like to be excluded in the workplace.
   f. I feel this experience will help me better recognize exclusive microbehaviors in my organization.
   g. If I saw this behavior in my organization, I would want to address it.
   h. This experience helped me recognize how subtle behaviors that exclude individuals can impact a person's motivation.
   i. I have witnessed or heard about a similar situation in my organization.
   j. I have experienced similar situations in my organization.
4. What is something that stood out to you from this experience?
5. [Measure of immersion within VR (Nowak & Biocca, 2003)] Please rate the experience on the following aspects on a scale from 1 (not at all) to 7 (very much).
   a. How involving was the experience?
   b. How intense was the experience?

c. To what extent did you feel immersed in the environment you saw/heard?
d. To what extent did you feel surrounded by the environment you saw/heard?

Post-Training Survey (taken after the last training session)

The following are answered on a scale of 1 (Strongly disagree) to 5 (Strongly agree):

1. I think I have high situational awareness when it comes to team dynamics.
2. I think I have high social and emotional intelligence when it comes to team dynamics.
3. I understand the value of increasing my social emotional competency.
4. I feel like I am pretty aware of the cultural nuances that take place in conversations on my team.
5. I feel comfortable to point out moments of exclusion I hear in conversations on my team.
6. My team culture supports and encourages difficult conversations.

Empathy

1. [IRI repeated from pre-survey] Please choose how well each of the following statements describes you in general on a scale from 1 (does not describe me at all) to 5 (describes me very well).

   a. I often have tender, concerned feelings for people less fortunate than me.
   b. I try to look at everybody's side of a disagreement before I make a decision.
   c. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (-)
   d. I sometimes try to understand my friends better by imagining how things look from their perspective.
   e. I would describe myself as a pretty soft-hearted person.
   f. If I'm sure I'm right about something, I don't waste time listening to other people's arguments. (-)
   g. When I'm upset at someone, I usually try to "put myself in their shoes" for a while.
   h. When I see someone being taken advantage of, I feel kind of protective toward them.
   i. I am often quite touched by things that I see happen.
   j. Before criticizing somebody, I try to imagine how I would feel if I were in their place.
   k. Sometimes I don't feel sorry for other people when they are having problems. (-)
   l. I believe there are two sides to every question and try to look at them both.
   m. Other people's misfortunes do not usually disturb me a great deal. (-)
   n. I sometimes find it difficult to see things from the "other person's" point of view. (-)

   1. In the past month, how often have you noticed subtle acts of exclusion in the workplace?
      - Often
      - Sometimes
      - Rarely
      - Never

9. One thing our team was able to address in our workplace culture through this engagement:
10. What is one company behavioral change that you think [the company] needs to make, to ensure a more trustful and supportive culture?
11. My biggest learning from this training:

Follow-Up Survey (taken nine weeks after last training session)

The following are answered on a scale of 1 (Strongly disagree) to 5 (Strongly agree):

1. I think I have high situational awareness when it comes to team dynamics.
2. I think I have high social and emotional intelligence when it comes to team dynamics.
3. I understand the value of increasing my social emotional competency.
4. I feel like I am pretty aware of the cultural nuances that take place in conversations on my team.
5. I feel comfortable to point out moments of exclusion I hear in conversations on my team.
6. My team culture supports and encourages difficult conversations.

Empathy

7. Please choose how well each of the following statements describes you in general on a scale from 1 (does not describe me at all) to 5 (describes me very well). [These items are taken from the Interpersonal Reactivity Index (IRI; Davis, 1980). They include the perspective taking and empathic concern subscales.]

    a. I often have tender, concerned feelings for people less fortunate than me.
    b. I try to look at everybody's side of a disagreement before I make a decision.
    c. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (-)
    d. I sometimes try to understand my friends better by imagining how things look from their perspective.
    e. I would describe myself as a pretty soft-hearted person.
    f. If I'm sure I'm right about something, I don't waste time listening to other people's arguments. (-)
    g. When I'm upset at someone, I usually try to "put myself in their shoes" for a while.
    h. When I see someone being taken advantage of, I feel kind of protective toward them.
    i. I am often quite touched by things that I see happen.
    j. Before criticizing somebody, I try to imagine how I would feel if I were in their place.
    k. Sometimes I don't feel sorry for other people when they are having problems. (-)
    l. I believe there are two sides to every question and try to look at them both.
    m. Other people's misfortunes do not usually disturb me a great deal. (-)
    n. I sometimes find it difficult to see things from the "other person's" point of view. (-)

8. In the past month, how often have you noticed subtle acts of exclusion in the workplace?
    - Often
    - Sometimes

- Rarely
- Never

9. When you witnessed these acts of exclusion, how did you respond?
    - I did nothing
    - Confronted the person who did the act of exclusion
    - Tried to smooth things over in the moment
    - Offered support to the person who may have been excluded, privately
    - Called out the behavior in a way that didn't single anyone out

10. Do you feel that your response was appropriate?
11. Why do you feel that your response was (or was not) appropriate?
12. If you did not respond, do you wish that you had? How do you wish you had responded?

13. In the past month, have you had any conversations about gender or racial bias?
    - Often
    - Sometimes
    - Rarely
    - Never

14. Within the last month, what specific actions did you take to make people around you feel more included? Please list them here.

15. One change I have seen in my team's workplace culture since going through this engagement is:
16. What is one company behavioral change that you think [the company] needs to make, to ensure a more trustful and supportive culture?
17. My biggest learning from this training:

Appendix B

Full Survey for Study 2: MEDTAPP

The questions in this survey are in reference to the virtual reality simulation project about social determinants of health in which you participated in either 2019 or 2020. This project will be referred to as "VRS" throughout the survey.

1. On what date did you participate in the virtual reality simulation project? (mm/dd/yyyy) If you do not recall the exact date, an approximate date is fine.

2. Have you participated in the online version of the simulation project? (y/n)

3. Likert scale – (1) strongly disagree to (5) strongly agree:
    a. As a result of the virtual reality simulation (VRS), my relationships with my patients have improved.
    b. Because of the VRS, I understand my patients better.
    c. The VRS made me more aware of how I interact with my patients.
    d. The VRS made me aware of social determinants of health that I did not previously know of.
    e. The VRS made me aware of implicit biases that I did not previously know of.
    f. Because of the VRS, I have more empathy for my patients.

4. Have you changed the way you approach your patients since the VRS? If so, how?

5. If you are able, please tell us a story about a patient interaction you think was more positive as a result of the VRS.

    a. How would this interaction have gone before the VRS and how did it go after the VRS?

6. Had you participated in other diversity/inclusion training(s) about social determinants of health before this one?

    a. If yes, how did the experience and impact of the VRS compare to the prior trainings you participated in?

7. Overall, what impact did the VRS have on you as a healthcare professional?

8. What is your job title?
    a. Clinical Psychologist
    b. Medical Resident
    c. Nurse
    d. Nurse Practitioner (or other APRN)
    e. Physician (indicate specialty: _____)
    f. Physician Assistant

g. Social Worker
h. Non-clinical Personnel
i. Other: _____

9. Approximately how much time do you spend with patients in a typical day?

10. What is your workplace setting?
    a. Federally Qualified Health Center
    b. Healthcare System Affiliated Clinic
    c. Hospital
    d. Private Practice
    e. Behavioral Health (i.e. community mental health)
    f. Other: _____

11. What is your sex?
    a. Male
    b. Female
    c. Other _____
    d. Prefer not to say