Sexual Harassment Bystander Training Effectiveness: Experimentally Comparing 2D Video to Virtual Reality Practice

Shannon L. Rawski, Ph.D. (Corresponding Author)
Organizational Behavior Group
Ivey Business School
Western University
London, Ontario
Email: srawski@ivey.ca

Joshua R. Foster, Ph.D.
Business Economics & Public Policy Group
Ivey Business School
Western University
London, Ontario
Email: jfoster@ivey.ca

Jeremy Bailenson, Ph.D.
Department of Communication
Stanford University
Stanford, CA
Email: bailenson@gmail.com

Data Availability Statement: The study data and some proprietary study materials (e.g., the training videos) are not publicly available due to restrictions from our Institutional Review Board Approval and the training provider. However, the authors can share deidentified study data to other researchers upon request. Our analyses are described in detail within the manuscript. Finally, the VR app used for the study is free and publicly available through the Apple App Store or Google Play Store. The name of the app is available to other researchers upon request.

Conflict of Interest Statement: We have no conflicts of interest.

Sexual Harassment Bystander Training Effectiveness: Experimentally Comparing 2D Video to Virtual Reality Practice

ABSTRACT

Sexual harassment (hereafter, SH) is a dysfunctional workplace behavior, resulting in negative outcomes for individuals and organizations. Since #MeToo, companies have been innovating to increase the effectiveness of SH training by incorporating new content (e.g., bystander intervention skills) and new technology (e.g., virtual reality, hereafter VR). However, research has yet to determine the best practices or the effectiveness of these new innovations. The current study hypothesizes that SH bystander intervention training will be more effective when VR practice scenarios are used rather than 2D video practice scenarios. We argue that the increased presence (i.e., the perception that people and places in a virtual simulation are real) afforded by VR should better replicate bystander experiences in real SH situations, thereby allowing trainees to develop bystander skills in a more realistic practice experience than 2D video provides. We experimentally test our hypothesis in a laboratory setting (N = 100). Our results show that the VR practice condition differed from the 2D video condition by increasing trainees' intentions to engage in indirect, non-confrontational, and widely applicable interventions (e.g., intervene by removing the target from the situation, approach the target to offer support later). However, our manipulation showed a negative effect on practice quantity (i.e., those in the VR condition explored fewer response options) and no effect on other operationalizations of training effectiveness (e.g., motivation to learn, knowledge, attitudes toward the training, intentions to directly confront the harasser, and intentions to formally report the harassment). Implications, limitations and directions for future research are discussed.

KEYWORDS: sexual harassment, bystander intervention, training, virtual reality, practice

Sexual Harassment Bystander Training Effectiveness: Experimentally Comparing 2D Video to Virtual Reality Practice

INTRODUCTION

In the wake of #MeToo, organizations have a renewed interest in preventing and managing sexual harassment (hereafter, SH) in the workplace. Researchers have identified two converging interventions for SH: bystander interventions (Bowes-Sperry & O'Leary-Kelly, 2005; Bowes-Sperry & Powell, 1999) and training interventions (Medeiros & Griffith, 2019). However, existing research on the topic of SH training has produced conflicting results on its effectiveness (Magley & Grossman, 2018; Roehling & Huang, 2018), culminating in a lack of clear, research-based best practices. There is even less research specifically on the effectiveness of SH bystander intervention training (Lee et al., 2019), which is focused on effectively executing social skills in a wide variety of nuanced SH incidents to stop harassment and protect targets. As such, SH bystander intervention training aims to develop *open skills* (i.e., applying general learning principles across many different situations, such that there is no single correct way to perform; Noe, 2017). For example, many SH, sexual assault, and street harassment bystander intervention training programs (e.g., Green Dot, Hollaback!, J.J. Keller & Associates, EVERFI) offer trainees a variety of potential interventions that can be used depending on the situation and the bystander's own strengths and weaknesses. These options may be categorized in a variety of ways to make use of memory aid acronyms (e.g., the 5 Ds for Hollaback!; Tam & Jin, 2021). For this study, we will refer to four types of bystander intervention response options, represented by the acronym IDEA, including (I) Intervene - causing a distraction or removing the target from the situation, (D) Direct - confronting the harasser and telling them to stop, (E) Elevate - formally reporting the harassment to an authority figure, (A) Approach - offering social support to the target after the incident.

Given the wide variety of potential SH situations (e.g., sexual coercion, gender harassment, and unwanted sexual attention; Fitzgerald et al., 1995), and intervention options (e.g. the four IDEA responses), we argue the stimulus generalization approach is appropriate to apply to bystander intervention training design. The stimulus generalization approach to training focuses trainees' attention on learning general principles and then demonstrating how those general principles can be applied in a variety of situations (Royer, 1979). One example of this approach may be found in customer service training, where employees must learn to address a wide variety of customer problems while adhering to general principles of active listening, empathizing, communication, and problem-solving within the constraints of company policy. The stimulus generalization approach is rooted in social learning theory (SLT; Bandura, 1982) and applies SLT's process of attention, retention, reproduction, and reinforcement. In particular, practice sessions during training offer trainees opportunities to behaviorally reproduce a learned skill and receive reinforcement in a wide variety of situations, enhancing skill development and transfer of new skills to the work context (Clark & Mayer, 2008).

Implementing practice opportunities in the context of SH bystander intervention training can be difficult for many organizations. Role plays may be one of the most obvious design options, but live role plays could potentially harass the employees acting in the victim role and hiring theater troupes to act out negative roles can be logistically and financially difficult to scale. Further, research on SH training has shown that trainees are sensitive to the social environment of training and can backlash against training if they perceive that the training is threatening their valued identities or social groups (Tinkler, 2012; Rawski, 2017). Many companies choose to implement computer-based training programs that offer scalable consistency in experience across all employees as well as opportunities to practice responses to a variety of scenarios, which may be

either video or text-based. Single-user computer-based training also offers trainees a safe place to make mistakes without facing ridicule or judgement from coworkers. However, computer-based training, even when it utilizes programmed instruction, does not fully capture the experience of witnessing SH firsthand and responding with a bystander intervention. There are important psychosocial factors (e.g., emotions, office politics) that will be absent from these types of computer-based practice scenarios.

Virtual Reality (hereafter, VR) training methods may offer an opportunity for increasing the immersiveness of SH bystander intervention training practice sessions, given VR's increased presence (i.e., the perception that people and places in a virtual simulation are real; Held and Durlach, 1992; Sheridan, 1992; Draper et al., 1998; Sanchez-Vives & Slater, 2005; Slater et al., 2006). Consistent with the Cognitive Affective Model of Immersive Learning (CAMIL; Makransky & Peterson, 2021), this attribute of VR represents an opportunity to further enhance trainees' acquisition and transfer of knowledge and skills, including the open skills needed for bystander intervention. Additionally, VR can more easily, cheaply, and safely place trainees in rare, stressful, or potentially harmful situations (such as conflicts involving SH) than in-person trainings and with more immersiveness than 2D computer-based trainings. Consequently, when designed as a single-user experience, VR offers all the dynamic benefits of SLT's reproduction and reinforcement phases of social learning while protecting trainees from taking on negative roles or making socially costly mistakes in front of their coworkers.

The current study seeks to determine the effects of single-user 360-degree immersive VR practice compared to single-user 2D video practice during SH bystander intervention training. We utilized a laboratory-based experimental design and randomly assigned 100 participants (university students and employees) to either experience a 2D video practice scenario or a 360-

degree immersive VR practice scenario following a SH bystander intervention training program. The content of the training and the practice scenario were kept constant with only the modality of the practice scenario differing across conditions. Results from our experimental study contribute to research on SH training, bystander intervention training, and VR training methods by isolating the effectiveness of VR practice methods compared to the more common 2D video methods.

LITERATURE REVIEW

SH Training

Traditionally, SH training has been defined as a systematic approach to increase employee learning related to (1) identifying and refraining from behaviors that constitute SH, and (2) following the organization's SH policy in reaction to the occurrence of SH (Alhejji et al., 2016; Goldberg, 2007). While the learning objectives of this type of training are laudable, reviews of the literature by both academics (Magley & Grossman, 2018; Roehling & Huang, 2018) and government agencies (EEOC, 2016) have found conflicting results, amounting to no clear best practices for SH training and little evidence to indicate its effectiveness. For instance, Bigham and Scherer (2001) found mixed effects, whereby SH training improved knowledge for all participants, but also increased male participants' victim-blaming and decreased their intentions to report harassment. Similarly, Goldberg (2007) found a mix of results whereby training increased targets' direct responses to harassers who gave them unwanted sexual attention, but did not have an effect on formal reporting behaviors. Further, there is some research that shows that traditional, legal compliance-based SH training programs can threaten employees' gender-based group dynamics and result in backlash effects such as resistance to organizational policy and increased intentions to engage in sexual behaviors at work (Tinkler, 2008, 2012; Rawski, 2017).

Given the lack of results demonstrated by traditional SH training, it is no surprise that researchers and practitioners have started innovating and designing new types of training programs to address SH in the workplace. Bystander intervention training has become a popular recommendation as an alternative style of training (Rawski & Workman-Stark, 2018; Lee et al. 2019). Theories on SH support this pivot toward bystander intervention training, suggesting that bystanders play a crucial role in mitigating the occurrence of SH (Bowes-Sperry & O'Leary-Kelly, 2005; Rawski, et al., *In Press*) and that a focus on the bystander role in training may be less threatening and more motivating to trainees than a focus on the harasser and victim roles (Rawski, 2017). However, research specifically on SH bystander intervention training for employees is still nascent, motivating calls for empirical studies on the topic (Lee et al. 2019).

VR Training Methods

In addition to the content and instructional strategy of SH training (e.g., traditional vs. bystander focused), the modality of SH training represents another design factor in need of additional research. Indeed, aside from Preusser et al. (2011) and Rawski et al. (2020), most empirical studies on SH training do not experimentally manipulate training modality, resulting in more questions than answers about how different technologies may causally influence SH training effectiveness. VR research on learning in environments similar to SH training reveals relevant differences among physiological and learning measurements across experimentally manipulated media. In one study, Slater (2013) demonstrates bystander intervention in potentially violent scenarios (e.g., a bar fight) is common in VR and increases when an in-group member is being targeted. Similarly, VR has been used to decrease implicit racial bias (Peck et al., 2013) and increase empathy regarding domestic violence (Seinfeld et al., 2018).

The direct and systematic experimentation of SH training media contributes to an ongoing debate over the role of modality in learning. In the view of McLuhan (1964) and McLuhan and Fiore (1967), it is argued that the characteristics of a communication medium shapes learning more than the instruction itself. Consequently, one would surmise the choice of learning modality is the single most important consideration to shaping learning outcomes. Conversely, arguments by Clark and Salomon (1986), Clark (1994), Sung and Mayer (2013) and Parong and Mayer (2018) suggest that media selections are incidental to learning, and only instruction methods are influential in certain core learning outcomes, such as achievement and motivation.

Since the conception of these theories, a large literature has begun to test their predictions, as well as several hybrid perspectives that lay between them. For instance, both Kozma (1994) and Makransky and Petersen (2021) offer an approach to understanding the roles of medium and instruction as co-deterministic, where the manipulation of one or both may cause variations in learning outcomes. One construct that seems to be central to understanding this medium-instruction interaction is the cognitive load demanded of the learner. Cognitive load is commonly measured as either intrinsic or extraneous, where the former relates to the inherent difficulty of the instruction material and the latter relates to how the instruction material is presented (Sweller, 1988; Paas et al., 2003; Reif, 2010). As the proposed interaction would suggest, there is evidence that cognitive load tradeoffs between various combinations of medium and instruction exist, which in turn influence learning outcomes. Specifically regarding the VR modality, VR has a demonstrated capacity to increase the learner's extraneous cognitive load relative to other media (Parong & Mayer, 2018; Makransky et al. 2019), an outcome that is attributed to its information-rich content. Under the assumption that intrinsic and extraneous cognitive loads are additive (Sweller & Chandler, 1994; Sweller, 2010; Mayer, 2017), this result suggests VR may impair

learning outcomes relative to other media. However, it has been shown this effect can be attenuated when learners are provided with learning materials related to the VR experience beforehand (a practice that is imposed in the present study's experimental design). One study (Meyer et al., 2019) manipulated pre-existing knowledge (pre-training vs. no pre-training) and instruction modality (2D video vs. immersive VR) and measured outcomes on knowledge retention. The results revealed that varying pre-existing knowledge did not cause changes to learning outcomes within the 2D video treatments, however the pre-existing knowledge aided in knowledge retention among VR users, suggesting it reduced extraneous cognitive load. Relatedly, it has been shown that immersive environments are important when learners need to recognize contextual nuances that can otherwise be hard to detect, particularly when there is a subjective threat present (Kroes et al., 2017). Consequently, when applied to SH training, VR environments may invoke psychophysiological and emotional responses among learners, which can be veridically important to their naturally occurring experience, thereby improving learning outcomes (Peck et al., 2013; Seinfeld et al., 2018). Moreover, given that a person's psychophysiological and emotional state has not traditionally been considered to be a relevant component of cognitive load, we argue that the intersection of SH training and VR technology allows for an important extension of this theory.

With two established and competing effects at play, this study relies on the Cognitive Affective Model of Immersive Learning (CAMIL; Makransky & Petersen, 2021) to guide comparative static predictions on learning outcomes across the forthcoming experimentally tested media. Specifically, CAMIL theorizes the immersive VR medium provides learners with relatively enhanced presence (cf. Sheridan, 1992; Riva, et al., 2003). Therefore, CAMIL predicts learners will experience greater situational interest that in turn will motivate sustained changes around intentions of future behavior. Indeed, immersive VR simulations stimulate similar

psychophysiological and behavioral responses as would real-life events (Coffey et al., 2017; Wang et al., 2012). As such, VR has been shown to be effective at motivating prosocial behaviors (Herrera & Bailenson, 2021) and developing social skills (Howard & Gutworth, 2020) such as public speaking (North et al. 2015), interviewing (Tsang & Man, 2013), and interpersonal communication (Morgan et al. 2014). These results are particularly applicable to the current study, given that the focus of bystander training is to develop social skills that result in a prosocial behavior (i.e., interventions to stop a SH incident).

A secondary benefit of VR's increased presence predicted by CAMIL is its effect on self-efficacy. Given the evidence on high-immersion feedback and its ability to increase a learner's sense of mastery (cf. Gegenfurtner et al. 2014; Makransky & Petersen, 2019; Makransky et al., 2020b; Johnson-Glenberg et al., 2021), CAMIL predicts immersive VR will provide learners with relatively greater expectations of self-efficacy. These benefits of VR should be especially pronounced in the context of SH bystander intervention training, which is focused on increasing bystander skills in a socially complex and risky situation (i.e., SH incidents). VR practice should provide trainees with the psychophysiological and emotional experience of encountering this social complexity and risk in an immersive, yet safe training environment, thereby increasing training effectiveness.

In measuring training effectiveness, this paper applies several well-studied metrics that capture the extent to which training achieves its intended objectives (Sitzman & Weinhardt, 2015), which can include reaction objectives (i.e., how trainees feel about the training program; Kirkpatrick, 1976), learning objectives (i.e., positive changes in trainees' knowledge, attitudes, or skills; Kirkpatrick, 1976), and/or performance objectives (i.e., transfer of learning to the work context; Holton, 1996). In addition, training motivation has been shown to be a key predictor of

these training outcomes (Colquitt et al., 2000), and training research often investigates trainees'

transfer motivation (i.e., trainee's intentions to exert effort toward using the training information

and skills in the workplace; Seyler et al., 1998) and other behavioral intentions (e.g., intentions to

report) as motivation- and self-efficacy-related proxies for transfer behaviors (Bingham & Scherer,

2001; Goldberg, 2007; Rawski & Conroy, 2020). Because the specific indicators of training

effectiveness are dependent on the specific type of training and its objectives, it is important for

any study on training effectiveness to define those specific indicators for the particular training

being studied. In the case of the SH bystander intervention training used as stimulus material in

this study (See Methods section), the training objectives included increasing knowledge and

behavioral intentions related to SH and bystander intervention techniques. Additionally,

motivation to learn, attitudes toward training (e.g., reactions), and practice quantity were also

investigated as more distal indicators of training effectiveness. As such, we predict the following:

Study Hypothesis: Relative to 2D video-based practice methods, 360-degree immersive VR practice methods following a SH bystander intervention training program will be associated with the following indicators of training effectiveness: (1) increased motivation to learn, (2) increased practice quantity, (3) decreased negative attitudes toward training, (4) increased training-related knowledge, and (5) increased intentions to engage in bystander interventions in response to future SH incidents.

METHODS

Research Ethics Approval

This research study was approved by the Institutional Review Board (IRB) for Research

on Human Subjects. Special modifications to the study's procedure (described in the Procedure

section) were made to keep participants and experimenters safe during the COVID-19 pandemic.

These modifications were also approved by the IRB.

Running head: SEXUAL HARASSMENT TRAINING VR PRACTICE

Participant Sample

The sample of participants consisted of 107 business students and employees recruited (see Procedure section for more information about recruitment) from a comprehensive public university in the mid-western United States. We removed seven participants from the analysis because of the following reasons: (1) one participant was removed for failing an attention check item (see Procedure section), (2) five participants were removed because they experienced technological problems (e.g., the VR app crashing[1]) during the lab study, and (3) one participant was removed because they reported feeling more than a slight amount of physical discomfort during the lab study.

Our final sample included 100 participants (50 per condition). The sample was 58% cisgender women, 40% cisgender men, and 2% non-binary. Additionally, 73% were students, 27% were employees, and 90% were white. The average age of the sample was 27.83 years (*SD = 11.50 years*). Participants had an average work experience of 11.02 years (*SD = 9.91 years*) and worked an average of 27.37 hours per week (*SD = 14.86 hours per week*). About 36% of the sample had experience managing employees, and 42% had experience supervising students. Forty-one percent of the sample had previously been a target of SH, 3% had been previously accused of sexual harassing someone else, 43% had previously been a bystander to SH, and 91% had previously participated in a SH training program. Thirty-five percent of the sample had previous experience using a VR headset.

---

[1]After an iPhone iOS update, the VR app was temporarily incompatible with iPhones. This unexpected problem was detected during a lab session, and prompted updates to the VR app so that it would be compatible with iPhones for future participants. The few participants who were affected by this technological problem, were unable to complete the study and thus, were removed from analysis.

To ensure random assignment we ran chi-square analyses (for categorical individual differences) and bivariate correlations (for continuous individual differences) to determine the statistical relationship between the assigned experimental condition and each of the aforementioned participant sample characteristics. None of the chi-square statistics nor the correlation coefficients were significant, suggesting that our assignment was indeed random across all individual demographic differences.

Stimulus Materials

The current experiment utilized commercially available stimulus materials, including two 2D training videos, a single-user 2D video-based practice scenario, and a free smartphone app used in conjunction with a Google Cardboard headset to view a single-user 360-degree immersive VR practice scenario. See Appendix A for screenshots and short descriptions of each training video, the practice scenario, and the VR interface. These stimulus materials were produced by a major US-based compliance training provider[2] as part of a commercially available anti-SH training program with a focus on bystander intervention. All videos (training and practice) were live action captures of professional actors depicting scenes in an office space.

2D Training Videos

All participants viewed both training videos on a computer screen. The first 2D training video depicted a female narrator presenting declarative knowledge about the definitions of SH and several examples of SH that occur in workplaces. This information was paired with video examples of different forms of SH, involving a diverse cast of actors in the various roles (e.g., harasser, target, bystander, manager). The second 2D training video depicted a male narrator presenting declarative knowledge about the four IDEA bystander response options, which

---

[2]For research purposes, the training app is available upon direct request to the corresponding author.

include (I) intervening in the ongoing harassment by causing a distraction or removing the target

from the situation, (D) directly confronting the harasser, (E) elevating the situation by reporting

to a manager, and (A) approaching the target to offer social support after the incident. The

training program used the acronym, IDEA, as a memory aid for trainees. The definitions of each

response were paired with tips for effectively executing the response option (e.g., several ideas

for how to intervene in ongoing harassment) and video examples of diverse actors (i.e.,

behavioral models) enacting the technique in response to a variety of SH situations.

Practice Modality Manipulation

In both conditions, participants were able to view (either in 2D video or in 360-degree

immersive VR video) an initial scene of a SH incident in which a man leers at and takes photos of

a woman without consent to which the woman clearly objects. Then, participants could select to

explore seven variations of the IDEA bystander responses. In the I, D, and A responses,

participants could choose whether to follow-up their initial response by choosing to either report

to the manager or not (i.e., using the E response as a follow-up to their initial response). Both

follow-up response choices, report and not report, would prompt another video to play, depicting

that choice. The E response (Elevate) initially reports to the manager, and so, no follow-up action

was offered to participants who chose the Elevate response. The IDEA response options were

offered in a vertical list in the order of the IDEA acronym. The stimulus materials did not allow

for randomizing the order of the options.

The SH scene and each of the IDEA response scenes were filmed from the point of view

of a bystander, and so, no visual depiction (e.g., actor, avatar) of the bystander was included in the

practice scenarios. The voice of the bystander in the IDEA response scenes was female in all

videos. After a response video played out in the practice scenario, participants had the option to

explore additional responses. Participants were informed that they could choose to explore as many response options as they wanted and that they could opt out of the practice scenario at any time. The length of the initial SH scene was 11 seconds, and the IDEA bystander response videos ranged in length from 11 to 51 seconds with an average length was 28.4 seconds. Participants were unable to see the length of videos before selecting their bystander responses to the initial scene.

We manipulated the modality of the practice scenario by randomly assigning participants to either experience a single-user 2D video practice scenario on a computer screen or a single-user 360-degree video practice scenario using a Google Cardboard VR viewer and their personal smartphone. The 2D video was derived from an optimal viewing orientation of the 360-degree VR video, and so, both conditions depicted the same exact performances of the SH and IDEA response scenes. This methodology effectively isolated the modality differences between the 2D video and VR practice conditions.

In the 2D video practice condition, participants viewed the SH incident video on a computer screen and then made selections about which IDEA bystander response option to view using their mouse. In the VR condition, participants did not utilize hand controllers (which are not included with a Google Cardboard headset), so choices were made using their gaze within the viewer to control a cursor (a small white circle that would expand into a larger white ring when hovered over a selectable button) and the gray button on top of the Google Cardboard headset to make selections about which IDEA bystander response option to explore. The cursor and navigation buttons were not visible while the immersive videos were playing, and only appeared after each video concluded. The Google cardboard did not include a head strap, so participants held the viewer up to their eyes with their hands.

The VR condition also required participants to download the required smartphone app (compatible with the Google Cardboard) and to watch a 10-minute instructional video about setting up their Google Cardboard and navigating the VR experience. Participants were able to pause this video to complete the setup sets (e.g., downloading the app, adjusting audio settings, removing their phone case, synching headphones, folding their cardboard viewer, aligning their phone in the viewer, etc.). Consequently, participants spent about 13 extra minutes on average (*SD = about 8 minutes*) in the VR condition compared to the 2D video condition (See Appendix B for an ANCOVA analysis including participation duration as a covariate).

Measures

Motivation to Learn. Motivation to learn was measured using nine items from Noe and Schmitt (1986). Responses were measured on a 5-point scale [1 = strongly disagree, 5 = strongly agree]. An example item is: "I will try to learn as much as I can from the SH training program." This measure exhibited an acceptable level of internal consistency [$\alpha = .85$].

Practice Quantity. Practice quantity was operationalized as a count of how many of the seven variations (See Stimulus Material section) of the IDEA bystander response options and follow-up report/not report responses a participant chose to explore during the practice scenario. Practice quantity ranged from zero to seven responses with an average of 2.63 responses explored [$SD = 1.76$].

Knowledge. We used ten items to measure SH and bystander intervention knowledge. These items were true or false declarative knowledge statements based on the content of the informational videos and the practice scenario included in the training program. Responses were measured on a 7-point scale [1 = definitely false, 2 = probably false, 3 = maybe false, 4 = I don't know, 5 = maybe true, 6 = probably true, and 7 = definitely true]. This scale allowed us to assess

not only the accuracy of a participant's knowledge, but also their degree of certainty. Example items [correct answer in brackets] include: "Sexual harassment does not include lewd sexual comments or jokes" [Definitely False] and "When you check in with the target of harassment to see if they are OK, you are using the Approach bystander response" [Definitely True]. Consistent with past research on training knowledge (Goldberg et al. 2019; Rawski & Conroy, 2020), our knowledge measure represents a formative construct because each item represents a unique piece of knowledge. As such, scores for this measure were calculated by taking the sum of items rather than the mean (Coltman et al. 2008), and no measure of internal consistency was calculated for the knowledge measure.

Negative (Backlash) Attitudes. Negative, or Backlash, attitudes against the SH training session were measured using eight items on a 5-point scale [1 = strongly disagree, 5 = strongly agree] from Rawski (2017). An example item is: "The scenarios discussed in this sexual harassment training session were ridiculous." The measure demonstrated acceptable levels of internal consistency [$\alpha = .83$].

Bystander (IDEA) Response Intentions. Bystander response intentions were measured using four items, one for each of the IDEA response options. Participants were asked, "How likely are you to use each of the following IDEA Anti-Harassment Action options in response to sexual harassment in the workplace?" followed by a list of each response option, including its acronym letter, its name, and a brief definition (e.g., E - Elevate the issue by reporting the harassment to a manager). Responses were measured on a 5-point scale (1 = very unlikely, 5 = very likely). Similar to our knowledge measure, these items represent a formative construct where each item represents a trainee's likelihood to use a different bystander response. So,

following Coltman and colleagues (2008), we took the sum of these four items rather than the mean and did not compute Cronbach's alpha.

Procedure

Participants were initially contacted through recruitment messages that were posted weekly to employee listservs and to business students' course websites. The recruitment messages advertised benefits of the research study, including the opportunity to improve SH training programs and to earn a free Google Cardboard VR viewer as a participation incentive. This message was carefully worded to avoid implying that the study involved using VR to avoid experimenter demand effects. Rather, the VR viewer was framed as a prize that participants would earn upon completion of the study. The cost of the VR viewer incentive is consistent with comparable cash payments offered to participants for similar research (Rawski et al., 2020). The recruitment messages directed potential participants to an intake survey (Survey 1) for the study.

Survey 1 screened potential participants based on willingness to participate in an in-person study during COVID-19, lack of COVID-19 symptoms and exposure, access to personal technological devices (e.g., smart phone, headphones, and laptop; See COVID-19 Precautions section), employment status (either currently employed or employed within the last 6 months), and low risk of motion sickness or vertigo. Those who were not willing or able to participate in-person (due to COVID-19), who did not have access to personal technology devices, who were not employed within the last 6 months, or who were at high risk for motion sickness or vertigo were screened out by Survey 1 and not invited to participate in the study.

For those who passed these screening questions, participants were asked to complete several pre-experiment questions and to schedule a time to come to the laboratory and complete the study in-person. The pre-experiment questions assessed participants' valence for SH outcomes

(e.g., the value for the ability to respond effectively to a SH situation), and their excitement to use a VR headset, as well as other technologies like laptops, smartphones, and headphones. Importantly, our recruitment and pre-experiment messaging did not indicate that anyone would use their VR headset during the study. Therefore, the most likely inference a participant might make regarding the study's interest in their excitement for using a VR headset would be related to the salience of its use as a recruitment incentive.

Upon arriving at the lab, all participants were screened again by the experimenter to ensure that they brought all of the required hardware devices (laptop, smartphone, and headphones). Then, the experimenter escorted participants to one of two identical rooms corresponding with their randomly assigned condition. Participants in different conditions completed the study in separate rooms to ensure that neither group's experience in the laboratory was influenced by observing the other. During the experimental sessions, all participants were observed by an experimenter and assigned their own work station where they were directed to set up their laptops and headphones. Regardless of condition, all participants were given a new Google Cardboard VR headset at their workstation to keep as an incentive for their participation. At this point, participants were directed to follow a URL to Survey 2. Survey 2 had five parts.

In Part 1 of Survey 2, participants read a description of their randomly assigned SH training condition. The training description informed participants that their SH training program consisted of two training videos and one practice scenario. The practice scenario was randomly assigned to be in either 2D video format on a computer or 360-degree video format in VR. After reading this description, participants answered four reading comprehension questions about their training description. If any of the comprehension questions were answered incorrectly, the participant was instructed to reread the training description and attempt the comprehension questions again. These

comprehension check questions were included to ensure that participants read and understood their randomly assigned training condition. Verification of this understanding was crucial to the validity of questions assessing participant's motivation to learn from their training condition (See next section).

In Part 2 of Survey 2, participants answered pre-training questions about their motivation to learn from their described training session. We included this pre-training assessment to determine if VR practice methods are associated with a motivational benefit compared to 2D video training methods. Next, in Part 3 of Survey 2, all participants viewed the same two training videos, which provided basic information about the definition of SH and the four IDEA bystander response options (see Stimulus Materials section and Appendix A). Then, in Part 4 of Survey 2, participants experienced their randomly assigned practice scenario, in either 2D video or 360-degree immersive VR (See Stimulus Materials section and Appendix A). Those in the VR condition watched an instructional video to set up their Google Cardboard viewer at this point in the procedure before beginning their practice session. Choices made during the practice session were recorded into Survey 2 (See Practice Quantity in the Measures Section). After choosing to stop their practice scenario, participants completed Part 5 of Survey 2 where they answered several post-training questions, including questions about presence, backlash attitudes against the training session, knowledge, and intentions to engage in the IDEA bystander responses if they observed SH in the future. This section of the survey also embedded an attention check question to verify that participants were reading the survey items carefully (See next section). Lastly, participants answered several demographic questions and one question assessing physical discomfort.

Comprehension Check, Attention Check, and Presence Manipulation Check

Comprehension Check. Given our interest in the post-manipulation, pre-training motivation to learn outcome, it was important to verify that participants understood the training they were assigned to experience after reading the training description at the start of the lab portion of the study. To verify that participants understood which type of training they would be experiencing, we included four comprehension check questions, each with four potential response options to disguise the study manipulation. The four comprehension questions included (1) What is the topic of today's training session? [Answer for both conditions: Sexual Harassment in the workplace], (2) How many informational videos will you watch in today's training session? [Answer for both conditions: 2], (3) In what format is the practice scenario in today's training session? [Answer dependent on condition: either Video-based or Virtual Reality-Based, respectively], and (4) in what format will you review the bystander response options to the practice scenario? [Answer dependent on condition: either Video-Based or Virtual Reality-Based, respectively]. If participants answered any of these questions incorrectly for their assigned condition, they were required to reread the training description and answer the questions correctly before progressing in the study.

Attention Check. We embedded one attention check item on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree) into Part 5 of Survey 2 amongst the outcome measurement items. The item read: "This is an attention check item. Please answer 'Strongly Agree' to this item to verify that you are carefully reading the survey." Those who did not answer this item correctly (n = 1) were excluded from analysis.

Presence Manipulation Check. Presence was included as a manipulation check. Past research shows that presence is greater in VR immersive video than in 2D video experiences

Running head: SEXUAL HARASSMENT TRAINING VR PRACTICE

(Hoffman et al., 2000; Slater, 2005), so we expect to find that our VR condition will also elicit greater perceptions of presence compared to our 2D video condition. We measured presence on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree) using 11 items adapted from Bailenson and Yee (2006). An example item includes: "During the practice scenario, I felt like the office setting was the real world." The internal consistency of these items was validated [α = *.87*].

COVID-19 Precautions

This experimental study was conducted during the COVID-19 pandemic, and as such, several precautions were taken to ensure the health and safety of all those involved in the study. The screening questions included in Survey 1, ensured that all participants were not experiencing COVID-19 symptoms before attending the lab session. This screening questionnaire also verified that all potential participants had access to required hardware (i.e., laptop, smartphone, and headphones) to participate in the study. Additionally, we gave all participants (regardless of condition) their own, new Google Cardboard VR viewer upon arrival to the lab study. By directing participants to use their own hardware devices and a personal VR viewer, we were able to avoid any potential cross-contamination of hardware among participants and reduce potential exposure to COVID-19. Workstations in the lab were sanitized before and after each participant's experimental session, and sessions were scheduled to allow for 2 meters of social distancing between workstations. All participants and experimenters were required to wear face masks in order to enter the building location of the lab and to participate in the study.

RESULTS

Means and standard deviations by condition can be found in Table 1. See Table 2 for the bivariate correlations between the study variables.

Insert Tables 1 and 2 about here.

We conducted a MANOVA to test the study hypothesis and determine whether VR practice methods are more effective than 2D video practice methods in terms of (1) motivation to learn, (2) practice quantity, (3) negative backlash attitudes toward training, (4) knowledge, and (5) bystander intervention intentions. MANOVA is a useful analysis in this context given our experimental data and the need to account for Type I errors and potential correlations among our dependent variables. Given this experiment's random assignment mechanism and highly controlled laboratory setting, we did not include any covariates in our main analysis used for hypothesis testing. However, we did conduct a post-hoc MANCOVA analysis to explore covariates that help to rule out alternative explanations of our results (See Appendix C).

See Table 3 for the MANOVA results of our main analysis. Consistent with the assumptions of MANOVA, Levene's test was not significant for all of our dependent variables, indicating that dependent variables had equal variances across the two experimental condition groups. Our Box's Test of equality of covariance matrices was also not significant [*Box's M = 31.99, F(21, 35324) = 1.42, NS*]. Our MANOVA results suggest that the experimental manipulation did have a significant effect on the multivariate outcomes [*Pillai's Trace = .35, p < .001; Wilk's Lambda = .65, p < .001*].

Insert Table 3 about here.

As a manipulation check we determined that the VR condition resulted in greater presence [*$\bar{x}$ = 3.96, SD = .59; F(1, 98) = 11.52, p < .001*] compared to the 2D video condition [*$\bar{x}$ = 3.56, SD = .59*]. This finding is consistent with past VR research (Hoffman et al., 2000; Slater, 2005) and indicated that our VR simulation was designed effectively.

The tests of between-subject effects revealed that the experimental condition did not have an effect on the post-manipulation, pre-training outcome, motivation to learn [$\bar{x}_{VR} = 4.06$, $SD_{VR} = .47$; $\bar{x}_{2D} = 4.10$, $SD_{2D} = .44$; $F(1, 98) = .17$, NS]. Therefore, knowing whether their training program would be 2D video-based or VR-based did not produce differential effects on participants' motivation to learn about SH and bystander intervention techniques. This result is not consistent with Part (1) of the study hypothesis.

Next, we observed unanticipated results for our in-training dependent variable, practice quantity. Those in the VR condition explored fewer IDEA bystander response options [$\bar{x} = 1.90$, $SD = 1.57$] during the practice scenario compared to those in the 2D video condition [$\bar{x} = 3.36$, $SD = 1.65$; $F(1, 98) = 20.56$, $p < .001$]. Translated into time spent in practice, we used the 11 second SH incident scene, the average length of the response scenes (28.4 seconds), and the average number of response videos watched in each condition, to determine that those in the 2D video condition spent about 1 minute and 47 seconds watching videos on a laptop during the practice session, while those in the VR condition spent about 1 minute and 5 seconds viewing 360-degree immersive videos in the VR practice session (See Appendix B for follow-up analyses including participation duration and physical discomfort as covariates). This result was not consistent with Part (2) of the study hypothesis.

To explore this result on practice quantity further, we conducted several follow-up chi-square analyses (See Table 4) to determine if our experimental condition had any effect on participants' decision to select particular IDEA response options, including their initial choices of I-Intervene, D-Direct, E-Elevate, or A-Approach, and their follow up choice to report or not report to a manger after each initial IDEA response choice (except for Elevate, since that response initially reports to a manager). Results indicate that participants were equally likely to initially

choose the Intervene response $[\chi^2 = 0.07, NS]$ and the Direct response $[\chi^2 = 1.03, NS]$ across

conditions. However, those in the VR condition were less likely to initially choose the Elevate

response $[\chi^2 = 4.11, p < .05]$ and the Approach response $[\chi^2 = 4.06, p < .05]$ compared to those

in the 2D Video Condition. Further investigation into the follow-up responses to either report or

not report to a manager show a clear pattern. Participants in both conditions were equally likely to

choose the Intervene $[\chi^2 = 2.77, NS]$, Direct $[\chi^2 = 0.00, NS]$, and Approach $[\chi^2 = 1.08, NS]$

responses with the no report follow-up choice. However, participants in the VR condition were

less likely to choose the Intervene $[\chi^2 = 31.56, p < .001]$, Direct $[\chi^2 = 9.01, p < .01]$, and

Approach $[\chi^2 = 5.00, p < .05]$ responses with the report follow-up choice when compared to those

in the 2D video condition. In combination with the significant results showing those in the VR

condition also initially chose the Elevate response less than those in the 2D video condition, there

is a clear pattern of those in the VR condition avoiding responses that report to the manager. This

aversion to experiencing responses related to reporting SH may have contributed to the lower

practice quantity observed in the VR condition.

Insert Table 4 about here.

Next, we examine the post-training outcomes. The MANOVA results showed no effect of

condition on backlash attitudes toward the training program $[F(1, 98) = 1.38, NS]$ or on knowledge

$[F(1, 98) = .46, NS]$, indicating a lack of support for Parts (3) and (4) of the study hypothesis.

However, we did observe that the VR condition resulted in significantly increased intentions to

use the IDEA bystander response options $[\bar{x} = 17.54, SD = 1.85; F(1, 98) = 4.95, p < .05]$ when

compared to the 2D video condition $[\bar{x} = 16.52, SD = 2.66]$. These results support Part (5) of the

study hypothesis and suggest that while participants learn and react to VR and 2D-video SH

bystander training programs equally well, they do develop greater intentions to transfer bystander knowledge and skills when VR practice methods are deployed in training.

To investigate the effect on bystander intervention intentions further, we ran the same MANOVA again (See Table 5) with the IDEA Bystander Response Intentions variable separated into its four distinct items: Intervene Intentions, Direct Intentions, Elevate Intentions, and Approach Intentions [*Pillai's Trace = .35, p < .001; Wilk's Lambda = .65, p < .001*]. Results from this analysis showed that the VR condition had positive effects on intentions to the Intervene and Approach responses [*Intervene Intentions: $\bar{x}_{VR}$ = 4.80, SD_{VR} = .40; $\bar{x}_{2D}$ = 4.58, SD_{2D} = .58; F(1, 98) = 4.90, p < .05; Approach Intentions: $\bar{x}_{VR}$ = 4.80, SD_{VR} = .40; $\bar{x}_{2D}$ = 4.48, SD_{2D} = .84; F(1, 98) = 5.91, p < .05*], but not the Direct or Elevate responses [*Direct Intentions: $\bar{x}_{VR}$ = 3.66, SD_{VR} = 1.06; $\bar{x}_{2D}$ = 3.36, SD_{2D} = 1.23; F(1, 98) = 1.71, NS; Elevate Intentions: $\bar{x}_{VR}$ = 4.28, SD_{VR} = .73; $\bar{x}_{2D}$ = 4.10, SD_{2D} = .89; F(1, 98) = 1.23, NS*].

Insert Table 5 about here.

DISCUSSION

The current study aimed to directly test the relative efficacy of a VR-based practice scenario against a 2D video-based practice scenario for SH bystander training in a controlled lab experiment. Results indicated that individuals are equally motivated to learn, accumulate equal levels of knowledge, and exhibit equal levels of backlash attitudes against SH training in both the VR and 2D video conditions.

However, those in the VR condition explored significantly fewer bystander response options during the practice session compared to the 2D video condition. This result is especially surprising given that our study accounted for differences in motivation to learn between the two conditions, finding none. Our follow-up chi-square analyses indicated that those in the VR

condition may have been avoiding bystander responses that involved reporting to the manager (i.e.,

the Elevate response and all follow-up reporting responses). This reporting avoidance may be due

to a combination of the increased presence of VR and the nature of the practice scenario itself. So,

perhaps, when immersed in the harassment scene, participants in the VR condition did not perceive

the harassment as severe enough to report. Alternatively, participants immersed in VR may have

had more fears about the social-psychological-emotional consequences of reporting, providing

some evidence that participants were deeply, emotionally engaged in the practice scenario. Future

research should further investigate the perceptual and emotional results of VR practice to parse

out these potential effects, especially as they may relate to increased cognitive load. Qualitative

interview methodology may be informative in determining the emotional work participants engage

in during a socially complex VR experience.

In addition, post-hoc follow-up analyses (See Appendix B) explored and statistically ruled

out two other potential explanations for this unexpected difference in practice quantity, including

the extra time spent downloading the app and setting up the Google Cardboard in the VR condition

and physical discomfort during the study. Though, our measure of physical discomfort occurred

at the very end of the study survey. So, it is possible that some participants did experience

discomfort in the VR condition causing them to explore fewer bystander response options during

practice, but that discomfort subsided by the time we measured this variable. Other potential

alternative explanations for the observed difference in practice quantity may include (1) the more

physically taxing navigation controls in the VR condition (i.e., controlling the cursor with gaze,

holding up the viewer), which may have led to practice fatigue and (2) greater learning efficiency

(Krokos et al., 2019), which may have reduced the time participants needed to master the training

content. This latter alternative explanation is supported by the observation that despite the

significant difference in practice quantity, knowledge was equivalent across the two conditions. Future research should further explore why VR may be associated with lower practice quantity compared to 2D video and consider utilizing VR headsets with head straps and hand controllers to reduce potential fatigue.

Finally, the experiment identified a positive effect of VR practice on intentions to engage in bystander interventions when compared to the 2D video practice condition. Specifically, we observed this effect of VR practice for the Intervene and Approach bystander responses, but not the Direct or Elevate responses. This pattern of results is notable in that the affected bystander response intentions represent informal and non-confrontational responses, whereas the unaffected response options represent formal and potentially confrontational responses. Additionally, the Intervene and Approach responses may be more generally applicable to all types of SH situations, whereas the Direct and Elevate responses may be more susceptible to situational or personal factors (e.g., severity of harassment, relationship with the harasser). Given that our measure of bystander intervention intentions referred to a general SH situation without including situational details, it is plausible the absence of such information could explain why the more situationally dependent response intentions were not affected by the experimental condition. Future research should determine if VR practice affects the type of bystander intervention chosen to respond to various SH scenarios with specific details included.

The observed pattern of significant effects on the Intervene and Approach responses may also be due to the increased presence of VR, which allows individuals to experience immersive video in a way that is closer to a real-life social interaction. Increased immersiveness of practice should also contribute to transfer of training according to both training research (van der Locht et al., 2013) and CAMIL theory (Makransky & Peterson, 2021). Future research could more

specifically test the mediating mechanisms that may be driving our observed effects in this study by using a more longitudinal design (e.g., measuring presence immediately after training and training outcomes after a time lag).

Implications

Our experimental study results contribute to both research and practice. First, our results contribute to the emerging literature on workplace SH bystander intervention training. While past research has recommended this type of training as a potentially more effective alternative to traditional, legal compliance-focused SH training (Rawski & Workman-Stark, 2018), little empirical research has documented the effectiveness of bystander-focused training programs (Lee et al. 2019). So, our study is a first step in exploring the many unanswered research questions about workplace SH bystander intervention training. Next, our research contributes to the immersive learning literature by empirically demonstrating the effectiveness of VR practice on a social skills-focused training program. CAMIL theory has suggested that the increased presence of VR may increase training effectiveness when this feature enhances instruction, but the increased cognitive load of VR may also decrease training effectiveness (Makransky & Peterson, 2021). Our results indicate that VR practice that follows the content of a 2D video training program increases informal, non-confrontational, and widely applicable bystander response options (i.e., Intervene and Approach responses) that represent lower social risk for bystanders while still supporting targets of harassment (Bowes-Sperry & O'Leary-Kelly, 2005). Consequently, the high presence of VR, perhaps even combined with the cognitive load of social factors within the practice scenario, may have represented a more true-to-life psychophysiological and emotional experience for trainees and motivated them to prefer less socially risky or more generally applicable interventions. Moreover, these results pose new questions to extend cognitive load theory. Namely,

do social risk assessments and emotional reactions to VR immersion increase cognitive load in a way that enhances or detracts from training effectiveness? Future research should seek to more directly answer this question. Finally, our results contribute to practice and help inform VR training designers and implementers of the effects of this training modality compared to 2D video training programs. Companies that seek to fully prepare employees to respond to SH in the context of complex social relationships and power hierarchies in work organizations will benefit from increases in informal, non-confrontational, and widely applicable intervention intentions resulting from training programs that utilize VR practice scenarios.

Limitations

While our controlled lab experiment made several contributions to research and practice, it also has certain limitations. We set out to compare VR practice to 2D video practice, and our results indicated where these two types of training practice modalities were equivalent for some types of training effectiveness indicators (i.e., motivation to learn, knowledge, backlash attitudes) and differentially affected others (i.e., practice quantity, bystander intervention intentions). However, future research should also include a true control condition (i.e., no practice, making sure to equate time spent on tasks across conditions) so as to determine the absolute effect of VR practice. We also note that this study was not able to measure the relative longitudinal effects of VR and 2D video practice, which would have provided evidence on knowledge retention, behavioral intentions and the recommended periodicity with which to offer these training sessions.

Another concern is the potential for social desirability motives to affect response choices in the context of SH training. While we did not directly measure social desirability, our use of random assignment to condition should have evenly distributed those with varying degrees of social desirability motives across our two experimental conditions. Additionally, we did assess

participants' valence for bystander actions, a potential indicator of social desirability, and included these valances as covariates in our follow-up MANCOVA analysis (See Appendix C), finding no effect on our results. Still, future research would benefit from a direct measurement of social desirability to rule out any potential effects it may have on SH training outcomes.

Additionally, several improvements could be made to our stimulus materials and VR equipment. First, future research should more carefully equate for time on task, noting that participants in our VR condition required additional time to set up and learn to navigate the practice experience compared to those in the 2D video condition. Better hardware (e.g., using hand controllers for navigation in VR instead of gaze) may also help reduce time differences between the two modalities.

Second, while our stimulus training material was designed to increase all types of bystander responses (i.e., Intervene, Direct, Elevate, and Approach), our results showed that the VR practice conditions only increased the responses of Intervene and Approach compared to the 2D video practice condition. Therefore, future research should determine how best to design VR training/practice that increases the salience of the social risks of intervening in a SH situation while still increasing more formal and potentially confrontational response options (i.e., Direct and Elevate), which also have a higher probability of stopping future harassment incidents (Bowes-Sperry & O'Leary-Kelly, 2005).

Third, our stimulus materials did not allow for the observation of unprompted bystander responses. In both conditions, participants were prompted to respond with a list of the IDEA response options. There also was not an option for participants to choose no response to the harassment situation (aside from quitting the practice session entirely). While these design choices were intended to encourage trainees to explore and practice the four IDEA response options, this

design also limited our ability to directly observe participants' unprimed behaviors in reaction to a SH scene. Consequently, we measured bystander intervention intentions as a proxy for bystander intervention behaviors. While research indicates that intentions are an antecedent of behavior (Rest, 1986), future research could improve on the design of our stimulus materials by allowing participants to more organically respond to a SH scenario without prompting and without constraining choices to specific types of responses. Another alternative would be to assess open-ended bystander responses to written scenarios.

Finally, our study utilized Google Cardboard VR headsets that doubled as research incentives. While this hardware choice presented definite benefits in terms of participants' health and safety during the COVID-19 pandemic (i.e., participants did not need to share headsets) and motivation to participate in the study, there were drawbacks to this hardware choice as well. First, the handheld headset may have contributed to practice fatigue. So, a headset with a head strap may be a better choice for future research. Additionally, by giving all participants a VR headset as an incentive, even those in the 2D condition, we may have disappointed participants who did not get to use their headset during the study. A different incentive structure, use of a shared headset sanitized with UV light, and measurement of participant emotions would avoid these issues in future studies.

CONCLUSION

In conclusion, VR practice after workplace SH bystander training is equivalent to 2D video practice on many important training effectiveness indicators (e.g., motivation to learn, knowledge, and backlash attitudes). Yet, VR's increased presence (compared to 2D video) likely results in our experimental study's observed differences in intentions to engage in bystander interventions, especially informal, non-confrontational, and widely applicable interventions (e.g., Intervene and

Approach responses). Organizations seeking to increase employees' enactment of bystander interventions should consider utilizing VR practice after a SH bystander intervention training program to increase its effectiveness.

REFERENCES

Alhejji, H., Garavan, T., Carbery, R., O'Brien, F., & McGuire, D. (2016). Diversity training programme outcomes: A systematic review. *Human Resource Development Quarterly*, *27*(1), 95-149.

Bailenson, J.N., & Yee, N. (2006). A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments, 15*(6).

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122.

Bingham, S. G., & Scherer, L. L. (2001). The unexpected effects of a sexual harassment educational program. *The Journal of Applied Behavioral Science, 37*(2), 125-153.

Bowes-Sperry, L., & O'Leary-Kelly, A. M. (2005). To act or not to act: The dilemma faced by sexual harassment observers. *Academy of Management Review, 30*(2), 288-306.

Bowes-Sperry, L., & Powell, G. N. (1999). Observers' reactions to social-sexual behavior at work: An ethical decision making perspective. *Journal of Management, 25*(6), 779-802.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21-29.

Clark, R. C., & Mayer, R. E. (2008). Learning by viewing versus learning by doing: Evidence-based guidelines for principled learning environments. *Performance Improvement, 47*(9), 5-13.

Clark, R. E., & Salomon, G. (1986). Media in teaching. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed.). Macmillan. New York.

Coffey, A.J., Kamhawi, R., Fishwick, P. et al. The efficacy of an immersive 3D virtual versus 2D web environment in intercultural sensitivity acquisition. *Education Tech Research Dev 65,* 455–479 (2017). https://doi.org/10.1007/s11423-017-9510-9

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*(5), 678.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, 61*(12), 1250-1262.

Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human factors, 40*(3), 354-375.

Equal Employment Opportunity Commission (EEOC). (June, 2016). *Select Task Force on the Study of Harassment in the Workplace*. Retrieved from https://www.eeoc.gov/eeoc/task_force/harassment/report.cfm#_Toc453686310

Fitzgerald, L. F., Gelfand, M. J., & Drasgow, F. (1995). Measuring sexual harassment: Theoretical and psychometric advances. *Basic and Applied Social Psychology, 17*(4), 425-445.

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology, 45*(6), 1097-1114.

Goldberg, C. B. (2007). The impact of training and conflict avoidance on responses to sexual harassment. *Psychology of Women Quarterly, 31*(1), 62-72.

Goldberg, C. B., Rawski, S. L., & Perry, E. L. (2019). The direct and indirect effects of organizational tolerance for sexual harassment on the effectiveness of sexual harassment investigation training for HR managers. *Human Resource Development Quarterly, 30*(1), 81-100.

Held, R. M., & Durlach, N. I. (1992). Telepresence. *Presence: Teleoperators & Virtual Environments, 1*(1), 109-112.

Herrera, F., & Bailenson, J. N. (2021). Virtual reality perspective-taking at scale: Effect of avatar representation, choice, and head movement on prosocial behaviors. *New Media & Society, 23*(8), 2189–2209. https://doi.org/10.1177/1461444821993121

Holton III, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly, 7*(1), 5-21.

Howard, M. C. & Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for social skill development. *Computers & Education, 144*. https://doi.org/10.1016/j.compedu.2019.103707

Johnson-Glenberg, M. C., Bartolomea, H., & Kalina, E. (2021). Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. *Journal of Computer Assisted Learning*.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and Development Handbook* (2nd ed.). New York: McGraw-Hill.

Running head: SEXUAL HARASSMENT TRAINING VR PRACTICE

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational technology research and development, 42*(2), 7-19.

Kroes, M. C., Dunsmoor, J. E., Mackey, W. E., McClay, M., & Phelps, E. A. (2017). Context conditioning in humans using commercially available immersive Virtual Reality. *Scientific Reports, 7*(1), 1-14.

Krokos, E., Plaisant, C., & Varshney, A. (2019). Virtual memory palaces: immersion aids recall. *Virtual Reality, 23*(1), 1-15.

Lee, S., Hanson, M., & Cheung, H. (2019). Incorporating bystander intervention into sexual harassment training. *Industrial and Organizational Psychology, 12*(1), 52-57. doi:10.1017/iop.2019.8

Magley, V. J., & Grossman, J. L. (2018). Do Sexual Harassment Prevention Trainings Really Work? Scientific *American Mind, 29*(2), 40–43.

Makransky, G., & Petersen, G. B. (2019). Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers & Education, 134*, 15-30.

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review,* 1-22.

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225-236.

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology.*

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning, 33*(5), 403-423.

McLuhan, M., & MCLUHAN, M. A. (1994). Understanding media: The extensions of man. MIT press.

McLuhan, M., & Fiore, Q. (1967). The medium is the message. New York, 123, 126-128.

Medeiros, K., & Griffith, J. (2019). # Ustoo: How IO psychologists can extend the conversation on sexual harassment and sexual assault through workplace training. *Industrial and Organizational Psychology, 12*(1), 1-19.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education, 140*, 103603.

Morgan, J. R. Price, M., Schmertz, S. K., Johnson, S. B., Masuda, A., Calamaras M., & Anderson, P. L. (2014) Cognitive processes as mediators of the relation between mindfulness and change in social anxiety symptoms following cognitive behavioral treatment, *Anxiety, Stress & Coping, 27*:3, 288-302, DOI: 10.1080/10615806.2013.839988

Noe, R. (2017). *Employee Training and Development*. New York: NY: McGraw Hill Education. ISBN: 978-0078112850

Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology, 39*(3), 497- 523.

North, M. M., North, S. M., & Coble, J. R. (2015). Virtual reality therapy: An effective treatment for the fear of public speaking. *International Journal of Virtual Reality, 3*(3): 1 – 6.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785.

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition, 22*(3), 779-787.

Preusser, M. K., Bartels, L. K., & Nordstrom, C. R. (2011). Sexual Harassment Training: Person versus Machine. *Public Personnel Management, 40*(1), 47–62.

Rawski, S. L., & Conroy, S. A. (2020). Beyond demographic identities and motivation to learn: the effect of organizational identification on diversity training outcomes. *Journal of Organizational Behavior, 41*(5), 461-478.

Rawski, S. L., & Workman-Stark, A. L. (2018). Masculinity contest cultures in policing organizations and recommendations for training interventions. *Journal of Social Issues, 74*(3), 607-627.

Rawski, S., Djurdjevic, E., Foster, J., & Soderberg, A. (2020). The Role of Trainer Characteristics in Intern Sexual Harassment Training Effectiveness. In *Academy of Management Proceedings* (Vol. 2020, No. 1, p. 19598). Briarcliff Manor, NY 10510: Academy of Management.

Rawski, S. L. (2017). Understanding employees' reactions to sexual harassment training: Interactional disruptions, identity threats, and negative training outcomes. (Doctoral dissertation). Retrieved from *ProQuest Dissertations & Theses Global*. (AAI10103180).

Rawski, S. L., O'Leary-Kelly, A. M., & Breaux-Soignet, D. (In Press). It's All Fun and Games Until Someone Gets Hurt: An Interactional Framing Theory of Work Social Sexual Behavior. *Academy of Management Review*, (ja).

Reif, F. (2010) *Applying Cognitive Science to Education. Thinking and Learning in Scientific and Other Complex Domains*. Cambridge, MA: The MIT Press.

Rest, J. (1986). Moral Development. New York: Praaeger.

Riva, G., Davide, F., & IJsselsteijn, W. A. (2003). Being there: The experience of presence in mediated environments. Being there: Concepts, effects and measurement of user presence in synthetic environments, 5.

Roehling, M. V., & Huang, J. (2018). Sexual harassment training effective- ness: An interdisciplinary review and call for research. *Journal of Organizational Behavior, 39*, 134-150.

Royer, J. M. (1979). Theories of the transfer of learning. *Educational Psychologist, 14*(1), 53-69.

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience, 6*(4), 332-339.

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L. E., Borland, D., ... & Sanchez-Vives, M. V. (2018). Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Scientific Reports, 8*(1), 1-11.

Seyler, D. L., Holton, E. F. III, Bates, R. A., Burnett, M. F., & Carvalho, M. A. (1998). Factors affecting motivation to transfer training. *International Journal of Training and Development, 2*(1), 16–16.

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments, 1*(1), 120-126.

Sitzmann, T., & Weinhardt, J. M. (2015). Training engagement theory: A multilevel perspective on the effectiveness of work-related training. *Journal of Management*, 1–25.

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., & Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PloS one, 8*(1), e52766.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., ... & Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PloS one, 1*(1), e39.

Sung, E., & Mayer, R. E. (2013). Online multimedia learning with mobile devices and desktop computers: An experimental test of Clark's methods-not-media hypothesis. *Computers in Human Behavior, 29*(3), 639-647.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review, 22*(2), 123-138.

Tam, R.. (Host). Jin, C. H. (Illustrator). (2021, May 20). COMIC: How to Intervene When Someone is Harassed or Attacked. [Audio podcast episode]. In Life Kit. *National Public Radio.* https://www.npr.org/2021/05/10/995423418/comic-how-to-intervene-when-someone-is-harassed-or-attacked

Tinkler, J. E. (2008). "People Are Too Quick to Take Offense": The Effects of Legal Information and Beliefs on Definitions of Sexual Harassment. *Law & Social Inquiry, 33*(2), 417-445.

Tinkler, J. E. (2012). Resisting the enforcement of sexual harassment law. *Law & Social Inquiry, 37*(1), 1-24.

Tsang, M. M. Y., & Man, D. W. K. (2013). A virtual reality-based vocational training system (VRVTS) for people with schizophrenia in vocational rehabilitation. *Schizophenia Research, 144*(1 – 3): 51 – 62.

van der Locht, M., van Dam, K., & Chiaburu, D. S. (2013). Getting the most of management training: the role of identical elements for training transfer. *Personnel Review*.

Wang, C.X., Calandra, B., Hibbard, S.T. et al. Learning effects of an experimental EFL program
in Second Life. *Education Tech Research Dev 60*, 943–961 (2012).
https://doi.org/10.1007/s11423-012-9259-0

Running head: SEXUAL HARASSMENT TRAINING VR PRACTICE

TABLE 1

Cronbach's Alphas, Means, Standard Deviations, and Effect Sizes

Table 1: Cronbach's Alphas, Means, Standard Deviations, and Effect Sizes

Variable	α	Grand Mean	SD	2D Practice Mean	SD	VR Practice Mean	SD	d
Post-Manipulation, Pre-Training								
Motivation to Learn	.85	4.08	.46	4.10	.44	4.06	.47	.53
In-Training								
Practice Quantity	-	2.63	1.76	3.36*	1.65	1.90*	1.57	1.61
Post-Training								
Knowledge	-	58.50	4.69	58.82	4.20	58.18	5.15	4.70
Backlash Attitudes	.83	2.08	.53	2.15	.53	2.02	.53	.53
Bystander Response Intentions (Total)	-	17.03	2.34	16.52*	2.66	17.54*	1.85	2.29
Bystander Intervene Response Intentions	-	4.69	.51	4.58*	.58	4.80*	.40	.50
Bystander Direct Response Intentions	-	3.51	1.15	3.36	1.23	3.66	1.06	1.15
Bystander Elevate Response Intentions	-	4.19	.81	4.10	.89	4.28	.73	.81
Bystander Approach Response Intentions	-	4.64	.67	4.48*	.84	4.80*	.40	.66
Presence	.87	3.76	.62	3.56*	.59	3.96*	.58	.59

N = 100
*Significant Mean Difference

TABLE 2

Bivariate Correlations

Table 2: Bivariate Correlations

Variable	1	2	3	4	5	6	7	8	9	10
1. Experimental Condition	-									
2. Motivation to Learn	-.04	-								
3. Practice Quantity	-.42***	.19	-							
4. Knowledge	-.07	.13	.19	-						
5. Backlash Attitudes	-.12	-.57***	-.24*	-.24*	-					
6. Bystander Response Intentions (Total)	.22*	.29**	.15	.00	-.36***	-				
7. Bystander Intervene Response Intentions	.22*	.27**	-.01	.14	-.36***	.64***	-			
8. Bystander Direct Response Intentions	.13	.17	.12	-.21*	-.13	.78***	.24*	-		
9. Bystander Elevate Response Intentions	.11	.33***	.23*	.23*	-.42***	.74***	.46***	.33***	-	
10. Bystander Approach Response Intentions	.24*	.14	.04	-.02	-.25*	.76***	.50***	.42***	.46***	-
11. Presence	.32***	.42***	.03	.16	-.44***	.39***	.31**	.17	.40***	.35***

N = 100

Experimental Condition: 0 = 2D Video Practice, 1 = VR Practice

* $p < .05$; ** $p < .01$; *** $p < .001$

TABLE 3

MANOVA Results

Table 3: MANOVA Results

Multivariate Tests of Experimental Condition				Tests of Between Subjects Effects of Experimental Condition						
Test	Value	$F_{(6, 93)}$	Part. Eta^2	Dependent Variable	SS	MS	MS$_{Error}$	$F_{(1, 98)}$	Part. Eta^2	d
Pillai's Trace	.35	8.31***	.35	Post-Manipulation, Pre-Training						
Wilks Lambda	.65	8.31***	.35	Motivation to Learn	.04	.04	.21	.17	.00	.00
				In-Training						
				Practice Quantity	53.29	53.29	2.59	20.56***	.17	.91
				Post-Training						
				Knowledge	10.24	10.24	22.07	.46	.01	.20
				Backlash Attitudes	.39	.39	.283	1.38	.01	.20
				Bystander Response Intentions (Total)	26.01	26.01	5.25	4.95*	.05	.46
				Presence	3.96	3.96	.34	11.52***	.11	.70

N = 100

* $p < .05$; ** $p < .01$; *** $p < .001$

TABLE 4

Bystander Response Choice Counts by Condition & Chi-Square Analyses

Table 4: Bystander Response Choice Counts by Condition & Chi-Square Analyses

Initial Response Choice	Count		χ^2
	2D Condition	VR Condition	
Intervene Response			
Explored	42	41	.07
Not Explored	8	9	
Direct Response			
Explored	23	18	1.03
Not Explored	27	32	
Elevate Response[3]			
Explored	18	9	4.11*
Not Explored	32	41	
Approach Response			
Explored	27	17	4.06*
Not Explored	23	33	
Follow-up Response Choice			
Intervene Response, No Follow-up Report			
Explored	42	35	2.77
Not Explored	8	15	
Intervene Response, Follow-up Report			
Explored	37	9	31.56***
Not Explored	13	41	
Direct Response, No Follow-up Report			
Explored	11	11	0.00
Not Explored	39	39	
Direct Response, Follow-up Report			
Explored	23	9	9.01**
Not Explored	27	41	
Approach Response, No Follow-up Report			
Explored	11	7	1.08
Not Explored	39	43	
Approach Response, Follow-up Report			
Explored	26	15	5.00*
Not Explored	24	35	

* $p < .05$; ** $p < .01$; *** $p < .001$

[3]The Elevate response initially reports to the manager, so there was no follow-up report response option after the initial Elevate response.

Running head: SEXUAL HARASSMENT TRAINING VR PRACTICE

TABLE 5

Follow-Up MANOVA Results

Table 4: Follow-up MANOVA Results

Multivariate Tests of Experimental Condition				Tests of Between Subjects Effects of Experimental Condition						
Test	Value	F(9, 90)	Part. Eta^2	Dependent Variable	SS	MS	MS$_{Error}$	F(1, 98)	Part. Eta^2	d
Pillai's Trace	.35	5.40***	.35	Post-Manipulation, Pre-Training						
Wilks Lambda	.65	5.40***	.35	Motivation to Learn	.04	.04	.21	.17	.00	.00
				In-Training						
				Practice Quantity	53.29	53.29	2.59	20.56***	.17	.91
				Post-Training						
				Knowledge	10.24	10.24	22.07	.46	.01	.20
				Backlash Attitudes	.39	.39	.283	1.38	.01	.20
				Bystander Intervene Response Intentions	1.21	1.21	.25	4.90*	.05	.46
				Bystander Direct Response Intentions	2.25	2.25	1.31	1.71	.02	.29
				Bystander Elevate Response Intentions	.81	.81	.66	1.23	.01	.20
				Bystander Approach Response Intentions	2.56	2.56	.43	5.91*	.06	.51
				Presence	3.96	3.96	.34	11.52***	.11	.70

N = 100

* p < .05; ** p < .01; *** p < .001

Running head: SEXUAL HARASSMENT TRAINING VR PRACTICE

TABLE A1

Post-Hoc ANCOVA Covariate Means, Standard Deviation, and Effect Sizes

**Table A1: Post-Hoc ANCOVA Covariate Means, Standard Deviations, and Effect Sizes**

Variable	Grand		2D Practice		VR Practice		d
	Mean	SD	Mean	SD	Mean	SD	
Participation Duration	36.17	9.23	29.66[*]	4.74	42.68[*]	7.96	6.55
Physical Discomfort	1.21	.41	1.20	.40	1.22	.42	.41

N = 100
[*]Significant Mean Difference

TABLE A2

Post-Hoc MANCOVA Covariate Means, Standard Deviation, and Effect Sizes

Table A2: Post-Hoc MANCOVA Covariate Means, Standard Deviations, and Effect Sizes

Variable	Grand		2D Practice		VR Practice		d
	Mean	SD	Mean	SD	Mean	SD	
Biological Sex	.58	.50	.60	.50	.56	.50	.50
Organizational Role	.73	.45	.72	.45	.74	.44	.45
Previous Victim Experience	.41	.49	.34	.48	.48	.51	.49
Previous Accused Harasser Experience	.03	.17	.04	.20	.02	.14	.17
Previous Bystander Experience	.43	.50	.48	.51	.38	.49	.50
Previous Sexual Harassment Training	.91	.29	.94	.24	.88	.33	.29
Previous VR Experience	.35	.48	.34	.48	.36	.49	.48
Valance of Protecting Others	4.49	.76	4.54	.73	4.44	.79	.76
Valance of Responding to Sexual Harassment	4.51	.67	4.52	.58	4.50	.76	.68
VR Excitement	3.49	1.19	3.54	1.20	3.44	1.18	1.19

N = 99 - 100

[*]Significant Mean Difference

APPENDIX A

Stimulus Materials

Training Video 1 Screen Shots & Description

Training Video 1 was 4 minutes and 31 seconds long and covered the definition of sexual harassment and several examples of SH. Both quid pro quo and hostile work environment were defined, and both types of SH were described as abuses of power and illegal sex discrimination under Title VII of the Civil Rights Act. The video described that SH can happen to or be perpetrated by both men and women, and by supervisors, coworkers, subordinates, third party contractors, or clients. Finally, the video described various forms of SH, including physical, verbal, non-verbal, pictorial, in-person, over the phone, via email or social media, on or off the worksite.

Running head: SEXUAL HARASSMENT TRAINING VR PRACTICE

Training Video 2 Screen Shots & Description
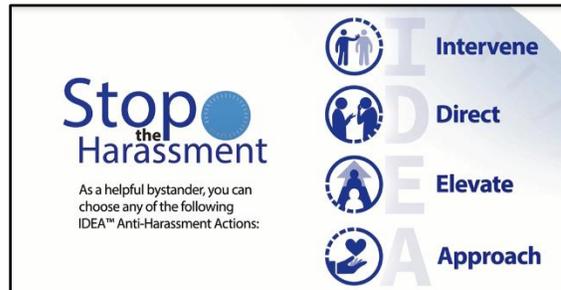
Training Video 2 defined and gave examples of the four IDEA Bystander response options, including Intervene, Direct, Elevate, and Approach. The video included a variety of harassment examples involving different actors of different genders, races, and ages in the target, harasser, manager, and bystander roles. Then, in each example, one of the IDEA Bystander response options was behaviorally modeled by a bystander in the situation. The video presented all four bystander response options as equally helpful to use in a SH situation.

Google Cardboard Set Up Video Screen Shots & Description

The Google Cardboard Set Up Video instructed participants in the VR practice condition on how to download the required App onto their phone, how to insert their phone into the Google Cardboard, and how to use the head set to navigate the VR Practice Experience. This video also included screenshots of the VR practice experience to better orient participants once they began using the headset.

Practice Session Screen Shots & Description

The practice session consisted of live action video examples of a SH incident. The incident depicted a white woman being leered at and photographed without consent by a white man in the coffee breakroom area of an office space. The dialog between the two characters implies that the behavior is part of a larger pattern of unwelcome sexualized behavior from man in the scene. After viewing the initial harassment scene, participants could choose to explore video examples of the IDEA bystander response options. Response options included: intervening to remove the target from the scene, directly telling the harasser to stop, elevating the issue to a manager (i.e., reporting), or approaching the target after the fact to offer support. After each initial response (except the elevate response), participants could also choose a follow-up response to either report or not report the harassment to a manager. Both the report and not report follow-up responses prompted an additional video to play.

Screen Shots of Virtual Reality Interface
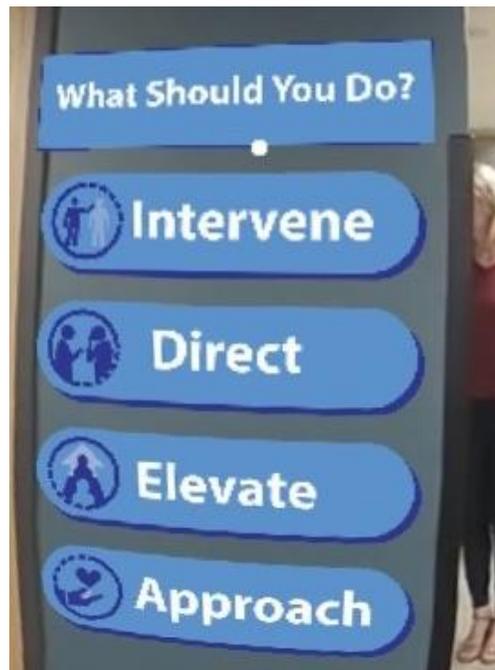
The virtual reality (VR) practice scenario was conducted through an app that participants downloaded onto their smart phones. Participants navigated the VR experience using the Google Cardboard headset (handheld, without a head strap). Participants were able to make choices within the practice session by using their gaze to move a cursor over blue buttons and then pressing the Google Cardboard's gray button (on top of the viewer) to make a selection.

Running head: SEXUAL HARASSMENT TRAINING VR PRACTICE

APPENDIX B

Post-Hoc ANCOVA with Participation Duration & Physical Discomfort Covariates

To explore alternative explanations for the surprising results related to practice quantity, whereby the VR condition was observed to have lower practice quantity, we ran a follow-up ANCOVA analysis for the practice quantity dependent variable, including participation duration and physical discomfort as covariates.

Measures

Participation Duration. Participation duration was measured using the time stamps for the start and end of Survey 2 (used during the lab study). Participants used the same survey to answer pre-training and post-training questions so these timestamps represent a fairly accurate measure of time spent in the lab portion of the study. To calculate duration, we subtracted participants' start time from their end time in minutes. We determined that participants took an average of 13.02 additional minutes (*SD = 7.96 minutes*) to complete the lab study in the VR condition compared to the 2D video condition. This difference was significant according to an independent samples t-Test with equal variances not assumed [*Levene's F = 9.803, p < .01; t = -9.939, p < .001*]. The additional time spent in the VR condition can be attributed to the need to download the VR app onto a smart phone, to watch a 10-minute, instructional video about how to set up the Google cardboard, and to navigate the experience with gaze (See Appendix A).

Physical Discomfort. Physical discomfort was measured with one item at the very end of Part 5 of Survey 2 (How much general physical discomfort do you feel right now?) on a 4-point scale that ranged from 1 = none at all to 4 = a severe amount. It should be noted that all those who indicated they felt more than a slight amount of general physical discomfort (a two on the measurement scale) were removed from analysis (n = 1). So, this measure accounts for the difference between those who felt no discomfort and those who felt slight discomfort.

Results

This follow-up analysis retained all previous data points from the original analysis (N = 100). See Table A1 for the covariate means and standard deviations.

Insert Table A1 about here.

When participation duration and physical discomfort were included as a covariates in a follow-up ANCOVA, our results still indicated that the experimental manipulation had a significant effect on practice quantity [$F(1, 96) = 42.19, p < .001$] with lower practice quantity being associated with the VR condition [$\bar{x} = 1.90, SD = 1.57$] and greater practice quantity being associated with the 2D video condition [$\bar{x} = 3.36, SD = 1.65$]. Physical discomfort was not significant in the model [$F(1, 96) = .82, NS$]. While participation duration did have a significant effect on practice quantity [$F(1, 96) = 17.76, p < .001$], it's inclusion as a covariate had no effect on the significance or direction of our observed results from our original analysis. So, these

alternative explanations do not sufficiently account for the decrease in practice quantity observed in the VR condition compared to the 2D video condition.

APPENDIX C

Post-Hoc MANCOVA Analysis with Individual Difference Covariates

We included various covariates in a follow-up, post-hoc MANCOVA analysis to rule out alternative explanations and inform future research. In particular, we controlled for a variety of relevant demographics, perceptions, and previous experiences.

Measures

Biological Sex. Biological sex was self-reported by participants during Part 5 of Survey 2. Response options included male, female, and intersex. Zero participants self-identified as intersexed, so we dummy coded the variable (0 = male, 1 = female).

Organizational Role. Given that our sample included both working adults and working business students, we measured organizational role at the university where we recruited our sample via self-report during Part 5 of Survey 2. We dummy coded participants' primary organizational roles (0 = employee, 1 = student).

Previous Experiences. During Part 5 of Survey 2, we assessed three types of self-reported previous SH experiences, including past experience as a victim, as an accused harasser, and as a bystander. Each of these three variables were dummy coded (0 = no previous experience, 1 = previous experience). Self-reported previous exposure to SH training and self-reported previous use of a VR headset were dummy coded as well (0 = no previous experience, 1 = previous experience).

Valence for Bystander Actions. On Survey 1, we asked participants how important achieving a series of work-related outcomes were to them. These outcomes included a variety of workplace achievements such as improving job performance and earning a pay raise. Embedded within this list were the items, "Protect others from workplace sexual harassment" and "Effectively respond to workplace sexual harassment". We kept these two items as separate one-item variables in the analysis. Responses were measured on a 5-point scale (1 = not at all important; 5 = extremely important). It should also be noted that these items also serve as a proxy for social desirability since participants may be prone to reporting that these altruistic actions are more important to them than they actually are.

VR Excitement. During Survey 1, participants were asked how excited they would be to use various technologies, including a laptop, a smart phone, headphones, and a virtual reality headset. The first three technologies were included so as not to reveal the study manipulation before participants arrived at the lab, but only the item pertaining to excitement to use VR was included in the analysis. Responses were measured on a 5-point scale (1 = not excited at all; 5 = very excited).

Results

Inclusion of biological sex, organizational role, previous experience as a target, an accused harasser, a bystander, with SH training, and with VR, valance for protecting others from

harassment and for responding effectively to harassment, and excitement to use VR reduced our sample size by 1 due to missing data (N = 99). See Table A2 for covariate means and standard deviations.

Insert Table A2 about here.

When the aforementioned individual difference variables were included as covariates in a follow-up MANCOVA, our results still indicate that the experimental manipulation had a significant effect on the multivariate outcomes [*Pillai's Trace = .35, p < .001; Wilk's Lambda = .65, p < .001*]. Of the included covariates, the only variables with significant multivariate effects were organizational role [*Pillai's Trace = .31, p < .01; Wilk's Lambda = .69, p < .001*] and valance for effectively responding to SH [*Pillai's Trace = .19, p < .01; Wilk's Lambda = .81, p < .001*]. Pillai's Trace and Wilk's Lambda were not significant for all other covariates.

Even with two covariates with significant multivariate effects, the pattern of results observed in the follow-up analysis remained consistent with our main analysis findings. There was no difference between the 2D video condition and the VR condition for motivation to learn [$F(1, 87) = .32$, *NS*], backlash attitudes [$F(1, 87) = 1.96$, *NS*], or knowledge [$F(1, 87) = .41$, *NS*]. Additionally, social presence was significantly higher [$\bar{x}_{VR} = 3.96$, $SD_{VR} = .58$; $\bar{x}_{2D} = 3.59$, $SD_{2D} = .57$; $F(1, 87) = 9.76$, $p < .01$], practice quantity was significantly lower [$\bar{x}_{VR} = 1.90$, $SD_{VR} = 1.57$; $\bar{x}_{2D} = 3.35$, $SD_{2D} = 1.67$; $F(1, 87) = 17.06$, $p < .001$], and bystander intervention intentions were significantly higher [$\bar{x}_{VR} = 17.54$, $SD_{VR} = 1.85$; $\bar{x}_{2D} = 16.51$, $SD_{2D} = 2.69$; $F(1, 87) = 4.40$, $p < .05$] in the VR condition compared to the 2D video condition. An additional MANCOVA analysis with the bystander intervention intentions separated into four distinct variables mirrored all previous results and again confirmed that the VR condition increased intentions to (I) Intervene [$\bar{x}_{VR} = 4.80$, $SD_{VR} = .40$; $\bar{x}_{2D} = 4.59$, $SD_{2D} = .57$; $F(1, 87) = 4.35$, $p < .05$] and (A) Approach [$\bar{x}_{VR} = 4.80$, $SD_{VR} = .40$; $\bar{x}_{2D} = 4.47$, $SD_{2D} = .84$; $F(1, 87) = 6.38$, $p < .05$], but not to (D) Direct [$F(1, 87) = 2.11$, *NS*] or (E) Elevate [$F(1, 87) = .39$, *NS*].

So, the included individual difference covariates did not alter the pattern or direction of significant results from our original analysis.