ABSTRACT

SAVILLE, JASON DONALD. Virtual Reality Learning Self-Efficacy: Does the Topic of VR Training Matter? (Under the direction of Dr. Lori Foster).

Virtual reality (VR) is increasingly being used for education and training, raising questions about how to assess the degree to which people feel competent learning through VR. In 2022, Saville and colleagues developed the Virtual Reality Learning Self-Efficacy Scale (VRLSE). The current experiment builds on their work. VR can be used for a wide range of training programs that vary with respect to the skill type and learning outcomes targeted. Whether the training content affects pre-training self-efficacy when VR is used for instructional purposes remains an open question. The present experiment tested whether the skill type and learning outcome targeted by a training program affect (a) how VRLSE scale items are interpreted; and (b) self-efficacy beliefs. A total of 831 participants recruited from the online research platform Prolific were randomly assigned to one of seven conditions resulting from a 2 (Skill Type: interpersonal or technical) x 3 (Learning Outcome: cognitive, skill-based, or affective) fully crossed factorial design plus a control group. Participants read about a hypothetical VR training program and provided ratings for VRLSE, potential moderators, and demographic items. Those in the control group read a generic vignette that did not specify skill type or learning outcome of the VR training. Those in the other six conditions read a vignette for VR first aid training highlighting one skill type and one learning outcome. Results confirmed prior work indicating the best-fitting model for the VRLSE scale is hierarchical, with one overall VRLSE factor and two subfactors of Learning and Interaction (Saville et al., 2022). Additionally, the VRLSE scale demonstrated measurement equivalence/invariance across all conditions. Finally, neither skill type nor learning outcome significantly affected the interpretation of VRLSE items or VRLSE beliefs themselves.

Virtual Reality Learning Self-Efficacy: Does the Topic of VR Training Matter?

by
Jason Donald Saville

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina

2023

APPROVED BY:

Dr. Lori Foster Committee Chair	Dr. Adam Meade
Dr. Kelly Lynn Mulvey	Dr. Paul Mulvey

DEDICATION

I dedicate this achievement to the community who taught me this:

Not seeing a path forward does not mean there is not a path. It just means you cannot see it.

BIOGRAPHY

Jason Saville grew up in Garner, North Carolina and made the long journey to Raleigh, North Carolina to attend North Carolina State University. Jason earned his Bachelor of Arts in Psychology in 2018 and went on to earn his PhD in Industrial-Organizational Psychology in the spring of 2023. While at NC State, Jason conducted research at the intersections between work, psychology, technology, and global development within the 4D Lab, under the steadfast guidance and mentoring of Dr. Lori Foster. Jason was also very fortunate to spend several years working as a graduate research assistant with the IntelliMedia Group within the Center for Educational Informatics, and it was during this time that Jason developed a love for interdisciplinary collaboration and research. Within teams of researchers including computer scientists, educators, digital artists, psychologists, and more, Jason led four scientific publications and was actively involved in six others. Examples of research projects he has participated with include prototyping first responder technology using virtual environments, identifying drivers of learner acceptance and use of virtual reality, investigating the relationships between self-efficacy beliefs and learning technologies, and exploring team communication behaviors within action teams. Additionally, Jason is very passionate about de-stigmatizing addiction and substance use disorders to better enable people to get the help and support they need. At this time, Jason works as an assessment consultant for Industrial/Organizational Solutions, collaborating with first-responder organizations across the United States to develop rigorous promotional processes for law enforcement and the fire service.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Lori Foster, for her steadfast support, guidance, and patience for longer than my graduate school career. It has been a pleasure learning under and sometimes alongside you over the years and I more than appreciate your commitment to help me make sense of what seems unintelligible: in my education, career, and often my own writing. Thank you for being a constant example of approaching life with grace, tact, and understanding.

I would also like to thank my other committee members who, by definition, I could not have done this without. Dr. Paul Mulvey, thank you for going unreasonably above and beyond for me and what seems like all your students; I have not met anyone as dedicated to the success of others as you, and I am very grateful to have experienced it. Dr. Adam Meade, thank you for somehow managing to have a compassionate and statistically sound answer to any question; your ability to communicate complicated topics in plain terms will be sorely missed. Dr. Kelly Lynn Mulvey, for a much shorter span of time, thank you for your dedication to open-mindedness and consideration for multiple viewpoints; yours is an example I hope to emulate as I navigate more of the world. To my temporary committee member Dr. Randall Spain, thank you for years of mentorship, fueling my passion for cross-disciplinary collaboration, and treating me as an equal.

Many more names come to mind when I think of my family and closest friends who assisted me in getting here and providing endless love and support along the way. Above all, I would like to thank the community I was introduced to at the lowest point in my life who have helped me build a life I never thought I would have. I experience joy, serenity, and freedom because of a decision (which I continue to make) to reach out and ask for help. For those who taught me gratitude is an action word, I hope my actions regularly demonstrate that gratitude.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Every year, the world's organizations spend hundreds of billions of dollars (USD) on training programs and initiatives (Bouchrika, 2020). Organizations around the globe seek to train individuals as efficiently and effectively as possible. There is no one-size-fits-all solution, and training designers may turn to a growing list of different and exciting technologies to help facilitate successful learning experiences. Virtual Reality (VR) is one such technology, which is growing in popularity (Alsop, 2022). Empirical attention is needed to fully understand how best to utilize VR for learning.

Learning technologies like VR are only one component leading to successful training. Individual characteristics, such as locus of control (Colquitt et al., 2000), conscientiousness (Colquitt et al., 2000), and motivation to learn (Goldstein & Ford, 2002; Noe, 1986; Quiñones, 1997; Salas et al., 2012) also play an important role in training effectiveness (Bell et al., 2017). An especially important individual characteristic is self-efficacy, which reflects the confidence individuals have that they can succeed at something. There is a rich literature establishing the importance of self-efficacy beliefs for training outcomes such as learning and job performance (e.g., Gist et al., 1991; Kirkpatrick, 1967; Noe, 1986). Additionally, several studies have established relationships between self-efficacy to use specific technologies, such as computers, and training outcomes such as learning engagement and learning performance (e.g., Chen, 2017; Compeau & Higgins, 1995; Gist et al., 1989). However, there is a dearth of research examining self-efficacy to use VR to learn. This is surprising, as the strengths and use cases of VR as a learning and training technology have been investigated for several decades.

Recently, a Virtual Reality Learning Self-Efficacy (VRLSE) scale was developed to help address this gap in the literature (Saville et al., 2022). As a nine-item assessment, the VRLSE scale stands as an instrument that could be deployed by organizations to assess employee readiness for VR training even before the training itself has been designed. The advantage of this early intervention is driven primarily by potentially significant time and resources saved: if organizations decide to use VR to train their employees, and if employees subsequently exhibit low confidence to learn using VR, employers can engage in targeted interventions to bolster those beliefs. Thus, a simple deployment of the VRLSE scale may have significant benefits for employers looking to maximize the outcomes of costly training programs.

The current study seeks to replicate and build upon Saville et al.'s (2022) currently unpublished VRLSE work. Recognizing the wide range of material that can be taught through VR, the present study examines individuals' self-efficacy to learn different types of material through VR training targeting a range of learning outcomes. Specifically, the purpose of this study is to: 1) confirm the previous factor structure of the Virtual Reality Learning Self-Efficacy scale using a new sample; 2) test whether the VRLSE scale structure remains the same when utilized for different training programs that vary with respect to skill type and learning outcome; and 3) experimentally examine whether virtual reality learning self-efficacy beliefs are affected by the skill type and learning outcome targeted by a VR-based training program.

Virtual Reality Training

Recent decades have seen training program designers shift from viewing trainees as passive recipients of information to viewing them as having an active role in their own learning (Bell et al., 2017). This shift has been facilitated by technology being woven into training that allows people to have more control over their learning (Noe et al., 2014). Rapid technological advances have allowed for exciting new possibilities for learning technologies to be incorporated into training, such as game-based learning (Graafland et al., 2019; Hendrix et al., 2016; Sousa &

Rocha, 2019; Spain et al., 2021), augmented reality (Barsom et al., 2016; Kaplan et al., 2021; Wang et al., 2017; Wang et al., 2018), and VR (Abich et al., 2021; Jensen & Konradsen, 2018; Lackey et al., 2016; Li et al., 2017).

VR is not a new technology by definition, nor is VR new to the training domain. However, increased commercial availability and affordability have allowed VR to be more accessible than in the past. Although there are many conceptualizations of what constitutes VR, I utilize the definition, “the sum of the hardware and software systems that seek to perfect an all-inclusive, sensory illusion of being present in another environment” (Biocca & Delaney, 1995, p. 63). As such, I refer specifically to technologies that make use of head-mounted displays to fully immerse a user in the virtual environment.

VR has been investigated in various learning domains of analytical problem-solving, interpersonal skills, procedural-practical knowledge, declarative knowledge, behavioral change, language learning, and others (Radianti et al., 2020). Research suggests VR is not equally effective at targeting all learning outcomes (see Jensen & Konradsen, 2018). For example, one study investigated cognitive learning outcomes of spatial awareness and factual information after learning about a historical site (Rasheed et al., 2015). Some participants were able to explore the historical site using VR, whereas another group of participants received classroom instruction about the site. After the learning experiences, all participants completed a written exam. The written exam revealed participants in the VR exploration condition scored higher on spatial awareness of the historical site, but lower on factual data items than participants in the classroom instruction condition (Rasheed et al., 2015). Another study examined the use of VR for interpersonal skill acquisition (Gutiérrez-Maldonado et al., 2015). Authors compared VR and desktop computer formats for teaching eating disorder diagnostic interview skills; no significant differences in skill acquisition were found (Gutiérrez-Maldonado et al., 2015).

Other research suggests VR can be beneficial when the outcome of interest is affective or involves evoking emotions (Jensen & Konradsen, 2018). For example, a randomized controlled trial was used to investigate VR exposure therapy for the treatment of social anxiety disorder (Anderson et al., 2013). Results indicated individuals who experienced VR exposure therapy significantly improved on clinical measures over and above untreated individuals on the waitlist. Additionally, when the same measures were used at a 12-month follow-up, participants who received VR exposure therapy demonstrated similar improvement compared to those who received active group therapy (Anderson et al., 2013).

In sum, research regarding VR's effect on learning outcomes is mixed. VR can be effective, but it is not always the best choice to maximize training outcomes. More work is needed to identify the conditions under which VR is more or less appropriate for training.

Self-Efficacy

The success of a training program depends not only on the technology utilized, but also on individual differences such as pre-training attitudes and beliefs. In other words, the attitudes and beliefs that individuals hold before training affect success during training. In particular, self-efficacy beliefs have been shown to play an important role in training effectiveness (Kirkpatrick, 1967; Noe, 1986). Self-efficacy, which has been studied at length (Bandura, 1977; Bandura, 1994/1998; Schunk, 1995; Stajkovic & Luthans, 1998; Talsma et al., 2018), refers to an individual's belief in their own ability to accomplish something (Bandura, 1997).

Self-efficacy theory was and is significantly influenced by social learning theory, which posits that behavior is often learned through observation, factoring in individual interpretation external to the observation itself (Bandura & Walters, 1977). The most famous examples of social learning theory in action are the Bobo doll experiments, also conducted by Bandura, in which children's decisions to imitate an adult's aggressive actions towards a doll were studied (Bandura et al., 1961). Drawing upon social learning theory, Bandura posited that self-efficacy is built primarily through 1) positive role modeling and 2) positive feedback and encouragement from others, as well as 3) individual mastery experiences and practice and 4) the interpretation of individual physical and emotional states (Bandura, 1977).

Self-efficacy beliefs are contextual to a specific topic or domain, as a given individual has different levels of confidence in different abilities in different contexts (Bandura, 2006). One such self-efficacy belief is training self-efficacy, or confidence in learning situations and learning new material (Noe, 1986). Training self-efficacy has long been considered an important antecedent of training performance (Gist, 1987; Latham, 1989; Tannenbaum et al, 1991). Research has highlighted individuals with high self-efficacy often outperform those with low self-efficacy (Gist et al., 1991). Additionally, self-efficacy can be better than past performance in predicting future performance (Gist et al., 1991). Although some training models suggest self-efficacy affects performance directly, other models have conceptualized training self-efficacy as primarily increasing motivation to learn, which in turn predicts performance (Noe, 1986). Specifically, individuals with higher training self-efficacy are often more motivated to learn information and behavioral skills than their counterparts with lower training self-efficacy. This increased motivation to learn results in more information being acquired and higher performance on skill-based assessments (Noe, 1986).

The theoretical association between training self-efficacy and motivation to learn has been empirically examined. In one study, subjects were told they would be participating in a training relating to an air defense task, and that their score on a pre-training knowledge test would determine if they were assigned to either a “remedial” or “advanced” training (Quinones, 1995). In actuality, participants were randomly assigned to one of two identical trainings after the knowledge test, irrespective of their scores. Before the training programs, participants rated their training self-efficacy and motivation to learn. Results revealed a significant relationship between condition assignment and subsequent training self-efficacy: individuals assigned to the remedial training condition exhibited significantly lower training self-efficacy than those assigned to the advanced training. Training self-efficacy increased motivation to learn, which in turn had a positive effect on learning, behavioral outcomes, and trainee reactions (Quinones, 1995).

One self-efficacy belief that is very similar to training self-efficacy is learning self-efficacy. Learning self-efficacy involves students’ appraisals of their capabilities to apply their knowledge to acquire new skills (Schunk, 1987). I did not find any prior work that compares and contrasts training self-efficacy and learning self-efficacy. The two variables are conceptualized, defined, and measured similarly. I will use the term “learning self-efficacy” moving forward to refer to both types of self-efficacy, as learning takes place in both educational institutions and the organizational training environment.

In addition to learning self-efficacy, other beliefs such as technology self-efficacy have demonstrable effects on elements of training effectiveness. Technology self-efficacy was originally conceptualized to represent an individual’s belief they can successfully utilize technology in general (Kass, 2014). However, given the domain specificity of self-efficacy, technology self-efficacy is most often measured by examining beliefs about specific technologies. Self-efficacy relating to specific technologies is what I refer to when I use the term technology self-efficacy. Technology self-efficacy scales have been developed to cover a range of areas including computer self-efficacy (Compeau & Higgins, 1995), internet self-efficacy (Jokisch et al., 2020), mobile technology self-efficacy (Menon et al., 2017), online technologies self-efficacy (Miltiadou & Yu, 2000), and information technology self-efficacy (Hwang, Lee, & Shin, 2016).

One of the most influential technology self-efficacy scales comes from Compeau and Higgins (1995), whereby the authors designed and validated a computer self-efficacy scale that captures individuals' beliefs concerning their abilities to competently use computers. Prior work has demonstrated computer self-efficacy is positively related to the decision to use computers (Hill et al., 1987), actual usage and enjoyment of computers (Compeau & Higgins, 1995), computer skill mastery (Gist et al., 1989), perceived ease of use of computers (Venkatesh & Davis, 1996), and learning performance and engagement (Chen, 2017). Additionally, computer self-efficacy has been found to be negatively related to computer anxiety (Compeau & Higgins, 1995; Wilfong, 2006) and computer anger (Wilfong, 2006). Therefore, in addition to the positive benefits of technology self-efficacy, strong technology self-efficacy beliefs may provide a buffer against negative interactions with technologies.

While measures have been developed to assess self-efficacy as it relates to a range of technologies, a VR self-efficacy measure has yet to be mainstreamed. However, as discussed next, there has been progress in this area upon which to build.

Virtual Reality Learning Self-Efficacy

Recently, Saville et al., (2022) created and validated the Virtual Reality Learning Self Efficacy (VRLSE) scale, defining VRLSE as the belief that oneself is capable of successfully using virtual reality technology to learn different topics. Saville et al.'s (2022) VRLSE scale was specific with respect to technology in that it is precisely focused on VR. However, it was designed to be generic with respect to training content – that is, it measured self-efficacy to learn any number of training topics through VR. Accordingly, the vignette preceding the VRLSE scale did not specify the content of the training program. Instead, the vignette asked participants to imagine they are in an educational/training program to learn a new skill, and that the program would use VR to teach them new things. Subsequently, respondents were asked to use a five-point Likert-type scale ranging from "Not at all confident" to "Totally confident" to rate their responses to items such as "How confident are you that you could use VR to learn and master difficult concepts without asking for help?"

Saville and colleagues utilized the online research platform Prolific (Prolific, 2023) to survey residents of the United States on 60 potential VRLSE items alongside measures of related constructs. Validation efforts used exploratory factor analysis to reduce the number of VRLSE scale items to nine. Confirmatory factor analysis for the resulting nine-item VRLSE scale revealed the best fitting model was hierarchical. Two first-order factors emerged, which the authors labeled Learning and Interaction. Learning and Interaction both load onto a second-order factor, VRLSE. For this reason, the VRLSE scale produces an overall score and/or two subfactor scores.

The Learning subfactor refers to beliefs individuals hold about successfully using VR to learn about a topic, whereas the Interaction subfactor refers to beliefs about successfully interacting with the VR system itself (Saville et al., 2022). A parallel can be drawn between the Learning subfactor and learning self-efficacy, as well as between the Interaction subfactor and technology self-efficacy. Additionally, Saville and colleagues (2022) established convergent and discriminant validity for the VRLSE scale by computing correlations between VRLSE and scores on several other measures, such as the Big Five Inventory version 4a (John et al., 1991). Results indicated the VRLSE scale demonstrated acceptable convergent and discriminant validity and was less associated with socially desirable responding than the other scales included in the research (Saville et al., 2022).

Skill Types and Learning Outcomes

The self-efficacy beliefs individuals hold before or during training do not exist in a vacuum. Training programs are not all created equally, and it is possible people feel more confident using VR in some training programs than in others. One way trainings may differ from one another is with respect to the skill type being taught. In addition, training programs also vary with respect to the learning outcomes targeted.

The term “skill type,” as used in this paper, refers to the learning domain itself. Training may be focused on hands-on “technical” skills or on relational “interpersonal” skills. Technical skills often involve working with data and things like equipment. Interpersonal skills often include communication and relating to others. The dichotomy of technical and interpersonal skills is found repeatedly in published research spanning a multitude of fields including but not limited to applied psychology (Lievens & Sackett, 2012), education (e.g., Bedwell et al., 2014), psychotherapy (e.g., Bennett-Levy et al., 2009), medicine (Fletcher et al., 2002; Yule et al., 2006), biology (Farrell et al., 2021), and information technology (Young, 1996). Some refer to technical skills as “hard skills,” and to interpersonal skills as “soft skills,” “non-technical skills,” “people skills,” “transferable skills,” “social skills,” or “generic competencies” (Matturro et al., 2019). A common trend among these conceptualizations is the emphasis that both technical and interpersonal skills are necessary for high performance in various professional roles and environments.

VR has been used to train technical skills, including but not limited to surgical techniques (Moorthy et al., 2003), construction (Osti et al., 2021), driving an automobile (Ropelato et al., 2017), and welding (Byrd et al., 2015). Additionally, VR has been used to train interpersonal skills such as job interview skills (Schmid Mast et al., 2018), public speaking (Anderson et al., 2005), and conversational skills for individuals with schizophrenia (Ku et al., 2007).

In addition to differences in skill types, VR training may also vary with respect to the learning outcomes targeted. Learning outcomes are addressed at length in the influential work of Kraiger, Ford, and Salas (1993). Drawing on the taxonomies developed by Bloom (1956) and Gagne (1984), as well as other prior theory and research, Kraiger and colleagues devised three categories of outcomes that make up learning: cognitive, skill-based, and affective. Cognitive learning outcomes include verbal knowledge, knowledge organization, and cognitive strategies (e.g., learning factual information about how bicycles work). Skill-based outcomes include compilation and automaticity (the processes through which skills are learned and mastered; e.g., applying knowledge and practicing how to ride a bicycle). Affective outcomes include both attitudes and motivations (e.g., motivation to practice riding a bicycle even after falling).

Knowing VR can be used to train different skill types and learning outcomes raises interesting questions about whether people feel more efficacious about learning some training content using VR than others. Self-efficacy to learn using VR should be examined in particular learning contexts that vary with respect to skill type and learning outcome. This is pertinent for organizations seeking to deploy the VRLSE scale and will indicate whether it is important to specify skill type and/or learning outcome before measuring people's VRLSE. To this end, Saville and colleagues (2022) suggested future implementations of the VRLSE scale be contextualized to specific learning domains.

Hypothesis and Research Questions

Addressing this suggestion, the present study explores if virtual reality learning self-efficacy is influenced by skill type and/or learning outcome. An important first step is to replicate the VRLSE scale structure found by Saville et al. (2022). Saville et al.'s (2022) results indicated an acceptable confirmatory model fit was achieved when the VRLSE scale was tested as a second-order hierarchical model. Thus, I predict the same model will successfully replicate when a similar vignette and items identical to those used by Saville et al. (2022) are provided to participants in a new sample. As in the previous study, I expect the second-order hierarchical model to fit better than one- and two-factor model solutions.

> **Hypothesis:** A hierarchical factor structure for the Virtual Reality Learning Self-Efficacy Scale consisting of two first-order factors and one second-order factor will fit significantly better than a one-factor or two-factor solution.

As noted, Saville et al.'s (2022) VRLSE scale assessed VR learning self-efficacy *generally*; it was agnostic to skill type and learning outcome. This is evidenced by the VR training vignette preceding the scale items. The vignette was generic in nature, describing no particular training topic. In addition to attempting replication with a generic VR training vignette, this study examines VRLSE for specific kinds of VR training. Rather than changing the scale items, the type of VR training under consideration can be specified in the vignette preceding the VRLSE scale, which can be written to prompt respondents to think about any number of VR training programs that vary with respect to the skill type and/or learning outcomes targeted.

Importantly, before testing whether skill type and learning outcome affect VRLSE beliefs, measurement equivalence must be established. Measurement equivalence/invariance (ME/I) analyses provide a way to test whether different groups of people interpret scale items identically (Putnick & Bornstein, 2016). For example, a common use of ME/I analyses is to ensure scales translated into different languages still function the same as the original validated measure (Schmitt & Kuljanin, 2008).

In the current context, ME/I is established if the VRLSE model hypothesized above emerges, regardless of the type of training under consideration. It is possible the skill type and/or learning outcome targeted by a training program could affect the way in which respondents interpret scale items. If this is the case, then the identified model solution for general VRLSE is not appropriate for all skill types and learning outcomes, and VRLSE scale scores cannot be directly compared across VR training programs.

In short, in addition to confirming the factor structure using a generic VR training vignette, this study also examines whether VRLSE scale items are interpreted the same by prospective trainees regardless of the VR training program's skill type and learning outcome.

> **Research Question 1:** Does the Virtual Reality Learning Self-Efficacy scale demonstrate measurement equivalence/invariance across different skill types and learning outcomes?

If ME/I is demonstrated, mean differences will be explored between groups to examine whether skill type and/or learning outcome affect the strength of VRLSE beliefs. At present, there is not sufficient research and theory to justify firm hypotheses regarding the effects of skill type and learning outcomes on VRLSE. Nevertheless, exploration of this topic is warranted. Given that self-efficacy beliefs are affected by the types of activities being appraised, the mode of capability being considered (e.g., cognitive, skill-based, or affective), and situational differences (Bandura, 2006), there are reasons to believe VRLSE could be affected by skill type and learning outcome.

For example, suppose some people have used VR in the past to learn technical skills. These people may develop strong confidence in their ability to use VR to learn technical skills, yet due to lack of exposure only have moderate confidence in their capacity to use VR to learn interpersonal skills. A finding such as this is plausible, as direct experience is one of the main sources of self-efficacy beliefs (Bandura, 1977). On the other hand, VRLSE could be higher for training targeting interpersonal rather than technical skills if trainees are especially prone to doubting their ability to learn to interact with data or things like equipment in the absence of "hands-on" experience.

Similarly, the learning outcome targeted by a training program could matter for VRLSE. For example, someone who believes VR to be poorly suited for the task of learning factual information may exhibit higher confidence to use VR for a behavior-based outcome, such as learning how to ride a bicycle. Additionally, individuals could experience either confidence or skepticism at the idea of VR being used for affective outcomes, such as producing motivation to continue practicing riding a bicycle. In short, there are a number of ways skill type and learning outcome could systematically result in different levels of VRLSE between participants. To date, this possibility remains unexplored.

Given the hierarchical nature of the VRLSE scale, VRLSE mean differences are examined with respect to individuals' beliefs about their capacity to (1) overall, learn through and interact with virtual reality (Overall scale), (2) learn through virtual reality (Learning subfactor), and (3) interact with virtual reality (Interaction subfactor).

Research Question 2: Does skill type (technical vs. interpersonal) affect the following aspects of virtual reality learning self-efficacy:

- **RQ2a:** Overall virtual reality learning self-efficacy?
- **RQ2b:** Beliefs about capacity to successfully learn through virtual reality?
- **RQ2c:** Beliefs about capacity to successfully interact with virtual reality?

Research Question 3: Does learning outcome (cognitive vs. skill-based vs. affective) affect the following aspects of virtual reality learning self-efficacy:

- **RQ3a:** Overall virtual reality learning self-efficacy?
- **RQ3b:** Beliefs about capacity to successfully learn through virtual reality?
- **RQ3c:** Beliefs about capacity to successfully interact with virtual reality?

Research Question 4: Do skill type and learning outcome interact to affect the following aspects of virtual reality learning self-efficacy:

RQ4a: Overall virtual reality learning self-efficacy?
RQ4b: Beliefs about capacity to successfully learn through virtual reality?
RQ4c: Beliefs about capacity to successfully interact with virtual reality?

CHAPTER 2: METHODS

Participants

Following recommendations by MacCallum and colleagues (1999) for determining necessary sample size for factor analysis, I examined scale communalities and the ratio of variables to factors within the VRLSE scale reported by Saville et al. (2022) and determined a group size of (*n* = 100) for each of the seven conditions would be sufficient for this study's purposes. An initial set of participants (*N* = 875) was recruited for this study via the online research platform Prolific (Prolific, 2023). Inclusion criteria required individuals provide consent, be 18 years of age or older, be residents of the United States, and not be students or employees of the university conducting this study. Participation in the study was estimated to take 8 minutes, but no minimum response time was required for payment to be rendered. Participants who submitted valid and complete data received $1.07 USD through Prolific, which corresponded with an hourly rate of $8.00 USD for the estimated completion time of 8 minutes. The median time for participants to complete all study materials was 4.53 minutes, corresponding to an hourly rate of $14.16 USD.

A total of (*n* = 44) individuals were excluded from data collection and/or data analysis for: not meeting participation criteria (*n* = 32), not completing 100% of survey items (*n* = 2), incorrectly responding to an instructed response item that read, "Please select 'A little confident,'" (*n* = 5), or answering "No" to a dichotomous item which asked participants whether their data should be used (*n* = 5; Meade & Craig, 2012). This resulted in a final dataset of (*N* = 831), which allowed for groups sizes to exceed the recommended (*n* = 100) for each of seven conditions. Additionally, the final dataset exceeded the necessary sample size indicated by a power analysis (G*Power; Faul et al., 2007) conducted for Research Questions 2-4 for an

ANOVA with fixed effects and interactions. Specifically, the power analysis suggested the sample size of (N = 831) would be able to detect a small-to-medium effect size of .1079 with a power of .80 at α = .05.

Participant ages ranged from 18 to 93, with an average age of 39.29 (SD = 14.33). When asked about which gender they most closely identify with, 47.3% of the final sample identified as female, 51.3% as male, and 1.4% as non-binary. Concerning ethnicity, 8.8% of participants identified as Hispanic or Latino/a/x and 91.2% of participants identified as not Hispanic or Latino/a/x. Concerning racial identity, 7.7% of participants identified as African, African American, or Black; 6.6% as Asian or Asian American; 79.3% as European, European American, or White; 0.1% as Native American or Alaskan Native; 0.2% as Native Hawaiian or Pacific Islander; and 5.4% as multi-racial. Additionally, 0.7% wrote in identities of Hispanic, Latino, Mestizo, Mexican/Chicano, and South American.

Regarding highest level of education attained, 0.7% of participants reported having completed less than high school, 1.0% some high school; 12.0% high school diploma or equivalent; 2.3% vocational training; 18.9% some college; 10.8% associate degree; 32.4% bachelor's degree; 3.7% some graduate school; 13.6% master's degree; 2.2% applied or professional doctorate degree; and 2.4% doctorate degree. When asked about current student status, 10.3% identified being full-time students, 4.6% as part-time students, and 85.1% as not being students. Regarding current employment status, 51.7% of participants identified being employed full-time (35 or more hours a week); 17.2% as employed part-time (less than 35 hours per week); 11.2% as not employed and looking for work; 12.9% as not employed and not looking for work; and 7.0% as retired.

Design

This study included seven conditions corresponding to seven VR training vignettes preceding the VRLSE scale. An overview of the skill type and learning outcome associated with each condition can be found in Table 1, and all vignettes can be found in Appendix A. The first condition represented general VRLSE, whereby a generic VR training vignette was administered to one participant group to enable Hypothesis 1's examination of the VRLSE scale's factor structure when skill type and learning outcome are not specified.

The generic VR training vignette used in Saville et al. (2022) for the VRLSE scale was modified in this study. In the prior study, the vignette referred to "an education/training program." In the present work, the generic VR training vignette refers simply to "a training program." Additional information defining immersive VR technology was also included in all vignettes to ensure all respondents understood this terminology. The remaining six conditions resulted from a 2 (skill type: interpersonal, technical) x 3 (learning outcome: cognitive, skill-based, affective) fully crossed experimental design. Manipulation of skill type and learning outcome occurred within the vignette preceding the VRLSE scale, with each vignette focused on one skill type and one learning outcome.

Skill type and learning outcome were manipulated within the common learning domain of first aid, where a variety of skills and learning outcomes can be taught and targeted. First aid includes important technical and interpersonal skills and can be applied outside of work as well as in the workplace across a variety of industries and organizations. First aid was selected as a domain because it is not one most people would consider themselves to be experts in. This is evidenced by a survey conducted by the American Heart Association (2017) which asked 2,003 workers across various industries in the U.S. about their first aid experience. Results of the survey indicated 33% of U.S. employees have never been trained in first aid or received cardiopulmonary resuscitation/automated external defibrillator (CPR/AED) training, yet more than 90% of those surveyed reported they would take first aid training if offered by their employer (American Heart Association, 2017).

Vignettes described VR training programs that teach either interpersonal or technical first aid skills and target learning outcomes that are either cognitive, skill-based, or affective, as classified by Kraiger, Ford, and Salas (1993). Virtual reality learning self-efficacy served as the dependent variable in this study.

Measures

Virtual Reality Learning Self-Efficacy

Saville et al.'s (2022) nine VRLSE items ($\alpha = .95$), presented in a random order, assessed participants' virtual reality learning self-efficacy. The item stem, "How confident are you that you could …" precedes the set of items. VRLSE has two subfactors: Learning (five items; $\alpha = .91$) and Interaction (four items; $\alpha = .93$). An example item from the Learning subfactor is, "... without any doubt, effectively use VR to learn about the program topic," and an example item from the Interaction subfactor is, "... easily learn how to operate VR controllers." Participants' VRLSE scores were calculated in three ways: as the average response value for all VRLSE items, the average response value for the Learning subfactor items, and the average response value for the Interaction subfactor items. Higher average values indicate higher levels of VRLSE. VR training vignettes for each participant group are shown in Appendix A. All VRLSE items can be found in Appendix B.

Saville and colleagues (2022) validated the VRLSE measure using a 5-point Likert-type scale, with response options of "Not at all confident" (1), "Slightly confident" (2), "Somewhat confident" (3), "Moderately confident" (4), and "Completely confident" (5). However, research has demonstrated that although numeric response options may have equal intervals between response options, the choice of words for written scale anchors may not be interpreted as equally spaced (Casper et al., 2020). Although the original VRLSE response options of "Slightly confident" (2) and "Moderately confident" (4) were selected by Saville et al. (2022) to represent 25% and 75% confidence respectively, the work of Casper and colleagues (2020) suggests individuals interpret these responses as 38% and 55%, respectively. As such, Saville et al.'s (2022) written anchors were replaced with anchors Casper et al. (2020) demonstrated to be more equally spaced. The new written anchors were, "Not at all confident" (1), "A little confident" (2), "Somewhat confident" (3), "Very confident" (4), and "Totally confident" (5).

Exposure to Virtual Reality

Three items were developed to assess participants' prior exposure to VR (α = .89; see Appendix B). These items asked participants to use a series of 5-point Likert-type response scales to self-report their current familiarity with VR, frequency of use of VR, and level of expertise with VR. Participants reported an average of 2.16 (*SD* = 0.94), on a 5-point scale where higher scores represent greater levels of familiarity, frequency of use, and expertise using VR.

Exposure to First Aid

Three items were developed to assess participants' prior exposure to first aid (α = .81; see Appendix B). These items asked participants to use a series of 5-point Likert-type scales to self-report their current familiarity with first aid, frequency of use of first aid, and level of expertise with first aid. Participants self-reported an average of 2.53 (*SD* = 0.85), on a 5-point scale where higher scores represent greater levels of familiarity, frequency of use, and expertise with first aid.

VR Sickness

One item asked participants to report the total number of hours of experience with using VR. Participants who provided a response greater than zero hours (n = 548, 65.94%), were asked, “Have you ever gotten physically sick while using VR?” This item used a 5-point Likert-type scale, with possible options of, “No” (1), “A little bit sick” (2), “Somewhat sick” (3), “Very sick” (4), or “Extremely sick” (5). The average response to this item was (M = 1.35, SD = 0.68), suggesting participants typically had not previously gotten sick from using VR, and those that did only experienced mild sickness.

Manipulation Checks

Two items shown in Appendix B were included to encourage participants’ awareness of the specific skill type and learning outcome (or lack thereof) present in their VR training vignette. Correct responses to both items were required to proceed. Instructions for the two items read, “Please answer the next two questions based on the training scenario above. The correct answer for each item will be required to ensure that you understand specific elements about the training before proceeding.” If an incorrect response was given for one or both items, the survey system indicated an incorrect response was chosen and asked participants to choose again until the correct answer was selected.

The first item assessed awareness of skill type. It read, “What topic will the training cover?” Possible responses included “No topic specified” (0), “Technical aspects of first aid” (1), and “Interpersonal aspects of first aid” (2). The second item, assessing awareness of learning outcome, read, “How will learning be assessed in the training?” Possible responses included, “No learning assessment specified” (0), “Multiple-choice knowledge test about first-aid terminology, principles, and theories" (1), "Role-play scenario demonstrating skills taught in the training" (2), and "Survey asking for ratings about attitudes" (3).

Attention and Data Quality Checks

One instructed response item assessed attention. This item, designed to mirror adjacent items and randomly presented amongst the VRLSE items, read "… please select 'A little confident,'" with possible response options of "Not at all confident" (1), "A little confident" (2), "Somewhat confident" (3), "Very confident" (4), and "Totally confident" (5).

One dichotomous item adapted from Meade and Craig (2012) assessed data quality. This question read, "Please consider how carefully you have responded to items in this study. In your honest opinion, should we use your data? You will be compensated regardless of your answer, but your honesty helps to improve our work." Response options were "No" (0) and "Yes" (1).

Procedure

Individuals self-selected to participate in this study using the online research platform Prolific (Prolific, 2023), which directed them to the online survey platform Qualtrics (Qualtrics, 2023). The survey platform could be accessed from a computer, tablet, or smartphone.

First, participants read an informed consent document that provided information on the study protocol and described the study as an investigation into attitudes toward different learning techniques. Those who wished to participate affirmed they met the inclusion criteria and provided electronic consent. After providing consent, participants were randomly and evenly assigned to view one of seven VR training vignettes for the Virtual Reality Learning Self-Efficacy scale (see Appendix A). Two manipulation checks were shown on the same page as the vignette, such that participants could review their assigned vignette while answering the manipulation check items.

The VR training vignette for general VRLSE was generic and did not provide participants with information on any specific skill type or learning outcome in the hypothetical training program. The other six vignettes described the focus of the training program as teaching either a technical or interpersonal skill type within the domain of first aid. Additionally, these six vignettes each gave an example of either a cognitive, skill-based, or affective learning outcome VR would be used to help achieve. To reinforce the targeted learning outcome, each vignette specified what kind of assessment would be administered at the end of training.

As shown in Appendix A, technical skill type was operationalized as a training program focused on the technical aspects of first aid. Interpersonal skill type was operationalized as a training program focused on the interpersonal aspects of first aid. Cognitive, skill-based, and affective learning outcomes associated with *technical* aspects of first aid were respectively described as: (1) learning information about technical aspects of first aid (e.g., the name and location of different parts of the body that are affected when someone is choking), (2) learning technical, hands-on first aid skills (e.g., strategies to assist someone who is choking, depending on characteristics of the situation), and (3) changing attitudes related to the technical aspects of first aid (e.g., boosting confidence for doing things such as applying any known technical first aid techniques when someone is choking).

Cognitive, skill-based, and affective learning outcomes associated with *interpersonal* aspects first aid were: (1) learning information about the interpersonal aspects of first aid (e.g., information such as why it is important to obtain consent for helping a person in distress, as well as the laws surrounding the consent process), (2) learning interpersonal first aid skills (e.g., how to obtain consent to provide care to someone in distress, including what to say and how to say it under different circumstances), and (3) changing attitudes related to interpersonal aspects of first aid (e.g., boosting confidence for doing things such as approaching someone in distress to ask for their consent to provide care).

After inputting the correct response to both manipulation check items, participants proceeded to the next survey page. Participants were informed they would review the same vignette they had just read, and that it would be repeated to enable them to reference the vignette while responding to new study instructions. On the following page, the same assigned vignette they already viewed was displayed again; in addition, participants were presented with instructions for responding to nine VRLSE items located on the same page. These items were displayed in a random order on one screen, along with the instructed response item. On the next screen, participants completed seven demographic items, presented in random order, assessing gender, ethnicity, race, age, highest level of education attained, student status, and employment status. Prior exposure to VR was assessed directly after the demographic items, and if participants reported more than 0 hours of experience with VR, they were asked one item assessing if they had ever gotten physically sick while using VR. First aid exposure was assessed following all VR items. The data quality check item appeared on its own page after all other items were presented. All items in the study, except for the instructed response item, required a response to proceed. Once all items were completed, participants were thanked and automatically directed back to Prolific.

CHAPTER 3: RESULTS

Confirming VRLSE Factor Structure

I hypothesized the greatest model fit for the VRLSE scale could be achieved via a hierarchical model containing one second-order factor of overall VRLSE and two first-order factors of Learning and Interaction. In this model, five items correspond to the Learning subfactor and four items correspond to the Interaction subfactor. This hypothesis was tested by conducting a series of confirmatory factor analyses (CFAs) utilizing data from the general VRLSE condition (n = 120). Fit statistics were assessed for each model, and qualitative consideration was given of multiple fit indices. While being cautious of relying exclusively on distinct cutoff scores, the guidelines used for good model fit included a non-significant Chi square test, SRMR values less than or equal to .08, RMSEA values less than or equal to .08, and CFI and TLI values each greater than or equal to .95 (Hu & Bentler, 1999; MacCallum et al., 1996).

The first CFA model tested was a forced two-factor solution, with five items corresponding to the Learning factor and four items corresponding to the Interaction factor. Results indicated the Chi-square test was significant and RMSEA was greater than .08, but other fit indices utilized were within recommended boundaries (see Table 2). As the Chi-square test is especially sensitive to Type I error as sample size increases, this was not an unexpected result. However, further investigation was given to the RMSEA value of .11, 90% CI [.07, .14]. Prior simulation research has indicated that RMSEA is often upwardly biased (Shi et al., 2019). Additionally, model simulations performed by Curran and colleagues (2003) indicate that when sample sizes are less than n = 200, the lower bound of the RMSEA confidence interval often exceeds the expected value, yet the expected value did not exceed the upper bound of the confidence interval in any model. Finally, investigations of the impact of fit statistic disagreement on model conclusions indicate that it is more concerning to have a “bad” CFI value paired with a “good” RMSEA value than vice versa (Lai & Green, 2016). Given these prior guidelines, I conclude that this model does demonstrate acceptable fit, which is consistent with the results of Saville et al. (2022). However, the model demonstrated a high covariance between the two factors ($\Sigma = .92$), which indicates a common factor may underlie the data. To investigate this, the second model tested was a forced one-factor solution (see Table 2).

A review of the fit statistics for the one-factor model appears to indicate a worse-fitting model than the two-factor solution. A likelihood ratio test (LRT) was conducted to statistically compare model fit between the two-factor and one-factor solutions (Vuong, 1989). To facilitate this, a two-factor model with factor covariance constrained to 1.0 was specified as the nested model to functionally represent a one-factor model, whereas a two-factor model where factors were free to co-vary was specified as the complex model. The chi-square test of differences for the LRT was significant, $\chi 2$ (19, n = 120) = 31.62, $p < .01$, indicating a two-factor solution provides significantly more information than a one-factor solution (Vuong, 1989).

With the high covariance between factors yet unresolved, the final model tested was the hypothesized hierarchical model, where five items were loaded onto the Learning factor, four items were loaded onto the Interaction factor, and both the Learning and Interaction first-order factors were loaded onto a second-order factor of VRLSE. The hierarchical model was specified using the marker method, whereby loadings of Learning and Interaction onto VRLSE were constrained to 1.0 and the loading of the first item from each subfactor was also constrained to 1.0. Fit statistics exhibited similar patterns as for the two-factor model. Specifically, the hierarchical model exhibited a significant Chi-square test and an RMSEA point estimate of .11, 90% CI [.08, .15] which lay above prior recommendations, while other fit statistics fell within recommended boundaries. For the same reasons as previously discussed for the two-factor model, the hierarchical model was deemed to demonstrate acceptable model fit (see Table 2).

Internal consistency was assessed via Cronbach's alpha for the Learning first-order factor (α = .91), Interaction first-order factor (α = .93), and overall scale of VRLSE (α = .95), with prior literature recommending Cronbach's alpha values greater than or equal to .70 (Nunnally, 1978). Cronbach's alpha values as high as these might otherwise warrant greater examination for redundancy within the scale. However, the limited number of items in each of the factors and overall scale do not suggest greatly inflated alpha values. Additionally, reducing the number of items in the measure to fewer than the nine items proposed by Saville et al. (2022) may result in lost variance. Taken together, these considerations and statistical analyses indicate sufficient evidence for a hierarchical VRLSE model solution outperforming one-factor or two-factor model solutions. This study's hypothesis is therefore supported.

Establishing Measurement Equivalence/Invariance

After confirming the factor structure and fit of the hypothesized model for the general VRLSE condition, RQ1 concerning ME/I across skill types and learning outcomes was investigated. To do so, ME/I was examined across all seven vignettes of the VRLSE scale, which differ across skill type and learning outcome and include the general VRLSE condition. ME/I analyses function iteratively by: (1) constraining model parameters within a measure's structural equation model, and (2) statistically comparing the fit of each iterated model against the fit of the previous model to determine if there is significantly different model fit. Model comparisons are assessed via a likelihood ratio test (LRT), with ME/I indicated by a non-significant Chi-square value. If a significant Chi-square value is found, the difference in CFI (ΔCFI) is used to assess if the change in model fit is meaningful. Past work has considered a change in CFI of greater than .01 (ΔCFIs > .01) to be a meaningful change in model fit for ME/I (Cheung & Rensvold, 2002).

Following suggestions from Vandenberg and Lance (2000), Schmitt and Kuljanin (2008), and Chen et al. (2005), ME/I was investigated in multiple ways: configural equivalence (equivalent factor structure across groups), first-order metric equivalence (equivalent factor loadings of items onto first-order factors), second-order metric equivalence (equivalent loadings of first-order factors onto VRLSE), first-order scalar equivalence (equivalent intercepts for all items), and second-order scalar equivalence (equivalent intercepts of first-order factors). For each subsequent model, the constraints of the previous model were retained. Although there are other forms of ME/I (namely residual equivalence and structural equivalence), equivalence at the configural, metric, and scalar levels is what is necessary to compare latent factor means or observed factor means (Schmitt & Kuljanin, 2008).

To this end, all seven participant groups were used in a series of multi-group confirmatory factor analyses (MG-CFAs), progressively testing each level of ME/I using an LRT to compare each new model to the previous model. Results of the ME/I models and tests can be found in Table 3. All fit statistics for each iterative model were within recommended boundaries except for significant Chi-square tests and RMSEA point estimates, and LRTs performed did not produce any significant Chi-square tests of differences. Most reasoning presented previously for acceptable model fit despite these two indicators are still applicable. However, prior work has indicated that less is known about the implications about disagreement between CFI and RMSEA within multi-group analyses (Lai & Green, 2016). Despite this consideration, the RMSEA confidence interval improvement moving from single- to multi-group analysis combined with other well-performing fit indices lead to the conclusion that each iterative model demonstrated acceptable fit. Research Question 1 was therefore answered in the affirmative: the VRLSE scale demonstrates ME/I across different skill types and learning outcomes.

VRLSE Mean Score Comparisons

Mean differences in VRLSE calculations between conditions were explored to address Research Questions 2-4, which concern whether skill type and/or learning outcome affect Learning, Interaction, and/or overall VRLSE. However, prior to examining mean scores of VRLSE between participant groups, prior exposure to VR and prior exposure to first aid were examined to determine whether they were confounded with this study's manipulation. Two one-way ANOVAs were conducted across the seven conditions to determine whether prior exposure to virtual reality or first aid significantly varied by condition. There were no significant differences between conditions for prior exposure to virtual reality [$F(6, 824) = .41$, $p = .87$] or prior exposure to first aid [$F(6,824) = 1.25$, $p = .28$].

Next, the means, standard deviations, and correlations of study variables were calculated (see Table 4), including three calculations of VRLSE scores. The results of the correlation analyses revealed moderate, positive, significant correlations between VR exposure and each of the three VRLSE calculations, indicating that individuals who had greater prior exposure to VR had higher VRLSE scores. Given this result, VR exposure was included as a covariate in the following three 2 (skill type) x 3 (learning outcome) factorial ANCOVAs.

A breakdown of each of the three VRLSE score averages for each condition can be found in Table 5. Averages ranged from a low of 3.29 (out of a possible 5) for the Learning score in the Technical and Affective condition to a high of 3.76 for the Interaction score in the Interpersonal and Affective condition. To investigate Research Questions 2-4, the main effects and interaction effects of skill type and learning outcomes on the VRLSE scores were examined.

Results revealed skill type does not significantly affect overall VRLSE (RQ2a; Table 6) and it does not significantly affect the learning (RQ2b; Table 7) or interaction (RQ2c; Table 8) aspects of VRLSE. Similarly, results revealed no significant effect of learning outcomes on overall VRLSE (RQ3a; Table 6), learning (RQ3b; Table 7), or interaction (RQ3c; Table 8). In addition, there was also no significant interaction effect between skill type and learning outcome on overall VRLSE (RQ4a; Table 6), learning (RQ4b; Table 7), or interaction (RQ4c; Table 8). In summary, neither skill type nor learning outcome significantly affected VRLSE scores before or after controlling for prior VR exposure.

Additional Analyses

Though not the focus of this study, several noteworthy relationships pertaining to gender and age emerged. As indicated by the significant correlations in Table 4, gender appears to play a role in VR-related attitudes and experiences. Specifically, gender was positively correlated with VR exposure. In other words, men were more likely to have prior experience with VR. Gender was also positively related to Learning, Interaction, and overall aspects of VRLSE, suggesting men had greater confidence in using VR for learning. To account for men's greater VR exposure, partial correlations were calculated. The significant correlations between gender and VRLSE persisted for Learning (r = .08, p < .05), Interaction (r = .18, p < .01), and overall VRLSE (r = .14, p < .01) when controlling for VR exposure, though the magnitude of each correlation was reduced from those shown in Table 4.

Age was negatively correlated with VR exposure (r = -.24, p < .05) and with Interaction (r = -.15, p < .05) and overall (r = -.11, p < .05) components of VRLSE. This suggests older individuals were less likely to have experience with VR and exhibited less confidence in interacting with VR technology. Partial correlations between age and VRLSE exposure revealed there was no significant correlation between age and Interaction or overall VRLSE after controlling for VR exposure. There was a significant and positive correlation between age and Learning ($r = .07$, $p < .05$) which did not exist prior to controlling for VR exposure, but this correlation was minimal. This means that age alone should not generally be used as an indicator of VRLSE. Instead, VR exposure may be a more meaningful variable.

CHAPTER 4: DISCUSSION

Self-efficacy beliefs shape important training outcomes such as learning and job performance (Gist et al., 1991; Kirkpatrick, 1967; Noe, 1986). Additionally, proactively managing self-efficacy beliefs can improve training experiences. However, self-efficacy beliefs are context-specific. Bandura (2006), who emphasized the importance of tailoring self-efficacy scales to specific contexts, noted: "One cannot be all things, which would require mastery of every realm of human life" (p. 1). As such, measures of self-efficacy have historically tapped into a wide range of topics, including the confidence individuals have to successfully use specific technologies. One technology that has not been widely studied in relation to self-efficacy is VR. As VR is increasingly being used to help achieve personal and organizational training goals, there is a need to better understand and measure self-efficacy to use VR.

The current study sought to explore VRLSE, which was first conceptualized by Saville and colleagues (2022) and best represented by a hierarchical model. Successful replication is an important step in gathering additional validity evidence for measurement scales. Thus, I predicted the VRLSE model structure established by Saville et al. (2022) would successfully replicate in a new sample. The data in this study supported the hypothesized model. As a result, VRLSE can be measured in three meaningful ways by calculating: 1) the average of the five Learning items, 2) the average of the four Interaction items, and 3) the average of all nine scale items. The complete VRLSE scale, including instructions for measurement and scoring, may be found in Appendix C.

This study also began to examine whether one's VRLSE changes depending on the skill type or learning outcome targeted by a training program. To examine this question, ME/I was first tested, with the data indicating neither skill type nor learning outcome significantly affect how individuals interpret scale items. In addition, analyzing mean score differences between conditions revealed VRLSE did not significantly vary across skill types and learning outcomes, even after controlling for prior exposure to VR. These results suggest one's sense of VRLSE may be stable across different training contexts.

Participants in this study, on average, demonstrated moderate confidence to successfully use VR to learn in their respective trainings. This was true whether VRLSE was calculated as Learning ($M = 3.39$, $SD = .86$), Interaction ($M = 3.65$, $SD = .91$), or overall VRLSE ($M = 3.51$, $SD = .84$). Additionally, the standard deviation values of the three VRLSE means indicate there is variation between individuals' VRLSE beliefs, perhaps driven by individual differences. This variation between individuals confirms this study's findings are not due to a floor effect, where most of the participants have very low VRLSE, or a ceiling effect, where most of the participants have very high VRLSE.

Though not the primary focus of this study, the data also yielded noteworthy findings concerning age, gender, and VRLSE. The significant negative correlation between age and VRLSE scores indicates older individuals have lower VRLSE than younger individuals; however, this appeared to be largely explained by differences in prior VR exposure. The significant correlation between gender and VRLSE scores indicates men are more likely than women to have higher levels of VRLSE. While the correlations between gender and VRLSE scores remained significant after accounting for VR exposure, the magnitude of each partial correlation was small ($r \leq .18$).

While there is no prior research exploring gender's relationship with VRLSE, there is a body of work concerning gender and technology self-efficacy. Specifically, past research suggests men are more likely to have higher levels of technology self-efficacy than women

(Busch, 1995; Whitley, 1997). However, research has also found it may be gender *roles* and not gender that drives the gap in self-efficacy. Huffman and colleagues (2013) surveyed university students on technology self-efficacy and the degree to which students related to items about traditionally masculine traits. Results of the work indicated men exhibited significantly higher technology self-efficacy than women, but this relationship was mediated by masculine gender roles over and above the influence of computer hassles or perceived structural support from the university (Huffman et al., 2013).

The rich literature on stereotype threat (Spencer et al., 2016; Steele & Aronson, 1995) also helps to elucidate gender and age disparities within this work. Stereotype threat exists when an individual, who is aware of a negative stereotype of a group they perceive themselves to belong to, is concerned about being judged negatively based on the stereotype (Spencer et al., 2016). Mental processes related to this awareness, such as active behavior monitoring to avoid stereotyped actions, drain cognitive resources that could otherwise be allocated for performance. This can lead to performance differences, or exacerbate existing differences, between groups that would not have been present to the same degree had the stereotype not been cued amongst stereotyped group members (Spencer et al., 2016).

In the context of the current work, the robust effects of stereotype threat could contribute to disparities in the confidence that women and older individuals feel to succeed in using VR for learning. Considering this alongside the correlation between gender and VRLSE, as well as between age and VRLSE, there is value in looking beyond simple demographics when predicting, measuring, or bolstering VRLSE. Variables such as gender roles, stereotype threat, and VR exposure may all contribute to apparent gender and age differences in VRLSE scores.

This study also found moderately strong correlations between VR exposure and all measurements of VRLSE. Perhaps a promising avenue for bolstering VRLSE is to increase individuals' exposure to VR. This is consistent with prior research identifying direct experiences, and especially enactive mastery experiences that provide individuals with knowledge and strategies for effective action, are strong sources of self-efficacy beliefs (Bandura, 1977; Bandura, 1997). However, it is also possible this relationship operates in the other direction. Individuals with higher VRLSE may be especially inclined to seek out VR experiences as compared to those with lower VRLSE. Future work should explore if this relationship is bidirectional to elucidate additional opportunities for successful training.

Theoretical Implications

The current work furthers self-efficacy theory (Bandura, 2006), which posits individuals do not feel equally confident across all situations, and confidence beliefs depend on contextual factors. This study showed skill type and learning outcome do not significantly affect learners' experience of VRLSE. This finding is of theoretical importance because it provides boundary conditions on the extent to which self-efficacy beliefs are contextualized. Put another way, there are limits on the degree to which context affects learners' perceptions of confidence in different environments.

The data provided in this study help us to better understand the nature of VRLSE, which can be conceptualized as a unitary construct, and at the finer-grained facet level. The hierarchical model of VRLSE has been demonstrated twice – first in an unpublished manuscript which developed and validated the VRLSE scale (Saville et al., 2022) and now in the current study. Findings suggest people's overall sense of VRLSE stems from a consideration of their ability to wield VR hardware as well as the ability to employ that hardware for learning purposes. The high correlation between Learning and Interaction within a two-factor model indicates that, while distinct, these two perceptions often coincide. As such, this study informs developing theory on VRLSE as a construct. With this statistical model and measure in place, we can begin to examine VRLSE more thoroughly as it relates to other constructs in its nomological network.

Practical Implications

The creation of a concise, nine-item measure of VRLSE presents opportunities for future VR trainings. Self-efficacy beliefs are known to influence training outcomes. Additionally, creating VR training with content that engages participants requires a great deal of time and money. The VRLSE scale may provide a low-cost solution to assessing individuals' readiness prior to training, potentially saving significant organizational resources associated with developing and deploying VR training.

The VRLSE scale can be used in two primary ways. First, it can be used to measure an individual's VRLSE *in general*, outside of the context of a specific training program. For example, a practitioner may want to know the workforce's VRLSE in general before deciding whether to develop a training using VR or a more conventional training method. If that practitioner knew their employees exhibited very low levels of VRLSE and there were limited opportunities for intervention, a more traditional training method may be a better choice. Alternatively, a researcher may wish to measure VRLSE in general for the purpose of evaluating an intervention's ability to bolster VRLSE, and/or to explore relationships with other constructs of interest. To measure VRLSE in general, users of the scale should administer the nine items with a VR training vignette that does not highlight specific learning topics.

Second, this scale can be used to measure how efficacious someone feels to use VR to complete a *specific* training program. For example, a trainer may wish to deploy the scale in advance of a training that has already been developed. In this instance, VRLSE could serve as an early predictor of training success, and the trainer could decide whether an intervention to boost VRLSE before training is necessary. To measure training-specific VRLSE, users of the scale should administer the nine items with a description of the specific training topic(s).

By examining VRLSE scores prior to training, either as VRLSE in general or as training-specific VRLSE, organizations can learn more about what aspects of VR training reduce learners' confidence. For example, perhaps a sample of potential trainees exhibits low scores on the Learning factor, yet high scores on the Interaction factor. This may suggest people are confident in interacting with VR technology but have less confidence in their ability to wield VR technology for learning purposes. In this instance, there may be something other than the technology itself that is causing doubts about the ability to master material in VR.

Alternatively, perhaps individuals exhibit strong Learning scores, yet low scores on Interaction. This information would be useful, as it suggests learners are not confident in their ability to wield the VR technology itself. Organizations could then focus on efforts to bolster confidence in VR as a platform, for example through direct exposure or education about how many different people achieve benefits or enjoyment from VR technology. From these examples, it is clear there is useful information to be gained through the measurement of Learning, Interaction, and overall VRLSE.

Limitations and Future Directions

The current work indicates that skill type and learning outcome do not significantly affect VRLSE beliefs. However, there are a few caveats to this finding. First, it is unclear whether these results would replicate if a different domain was chosen. First aid was selected as the domain for this study because it includes both technical and interpersonal aspects that are applicable across industries, and because most people are not experts in first aid. In this study, participants possessed a wide range of experience with first aid, indicated by the composite variable of first aid exposure. Participants also did not have significantly different levels of first aid exposure across conditions. Analyses did not identify any differences in model fit or VRLSE scores between the generic VR training vignette condition and other conditions, despite the generic condition not mentioning first aid at all. Additionally, while the original study of VRLSE did not investigate measurement equivalence or between-group comparisons, the work of Saville and colleagues (2022) did establish the hierarchical model of VRLSE acceptably fit the data when the vignette was purposefully generic. While each of these considerations suggests that the skill type and learning outcome results of this study are not unique to the domain of first aid, future research outside the domain of first aid would be useful.

Second, this study makes no claims about the degree to which any trends identified in this work do or do not exist for training mediums other than VR. For example, the finding that skill type and learning outcome do not significantly affect VRLSE attitudes may or may not be consistent with how skill type and learning outcome affect self-efficacy beliefs in non-VR-based training – an unexamined topic worthy of investigation in its own right. Instead, the current work is focused exclusively on individuals' appraisal of VR trainings.

Another limitation is participants were not presented with a real training using VR. Further, 34% of participants (*n* = 283) did not have any prior experience with VR. These are noteworthy limitations, as technology self-efficacy is sometimes assessed after individuals have an experience with the technology. However, self-efficacy beliefs do not require prior experience. We as individuals may hold beliefs about our ability to succeed at utilizing different technologies or participating in various tasks or activities without having prior experience with those things. For example, someone who has never been employed likely has beliefs, positive or negative, about their confidence in being able to succeed in their first role. Additionally, someone who has never used a new handheld tablet may have high confidence in successfully using the tablet based on their experience with other handheld technologies. Finally, perhaps someone who has a long history of using VR for gaming purposes believes they would excel at adapting to a VR system (i.e., the Interaction factor), but has low confidence they would be able to learn using VR (i.e., the Learning factor). As such, it is important for organizations implementing VR training to consider that beliefs are being formed by VR veterans and non-users alike, and these beliefs may impact the success of the training.

This study approaches self-efficacy from the lens of Bandura's work, in which self-efficacy beliefs are specific, nuanced, and contextual in nature. However, there is significant research that has been performed on the variable of generalized self-efficacy. For example, the work of Judge and colleagues have investigated the trait of core-self evaluations, which integrates self-esteem, locus of control, generalized self-efficacy, and low neuroticism (Judge, 2009; Judge & Bono, 2001). Meta-analytic work has typically found generalized self-efficacy to perform well as a predictor of job performance, especially in concert with the remaining facets of core self-evaluations (Judge & Bono, 2001). While generalized and specific self-efficacy beliefs do not inherently exist at odds with one another, there are additional considerations for practical application of contrary theoretical approaches.

One potential next step for this work includes having VRLSE, along with generalized self-efficacy and other various other constructs that comprise self-efficacy's nomological network, measured within an organizational context where VR is used to deliver training. Future work could examine how VRLSE relates to outcomes such as amount of information learned, learning engagement, and job performance by deploying the VRLSE scale for individuals who will go on to participate in a real organizational training. In particular, future work could also investigate whether VRLSE has predictive ability over and above that of generalized self-efficacy for organizational outcomes. This applied investigation of VRLSE would assist in identifying the role of VRLSE in predicting outcomes of interest for organizations, as well as increase the defensibility of VRLSE both as a construct and as a measure.

Additionally, future work should examine the interplay between Learning and Interaction components of VRLSE. It is possible Interaction provides a ceiling effect for beliefs to successfully wield VR technology for Learning. Perhaps an individual is not able to feel very confident in their ability to use VR to learn if they are not also very confident in interacting with VR equipment. Future work examining the relationship between Learning and Interaction may shed light on whether interventions designed to bolster Interaction scores will also boost Learning scores as well.

It is also possible that the extent to which VRLSE beliefs can be strengthened is dependent on whether an individual believes they have the capability to improve their own abilities – the extent to which the individual has a “growth” or “fixed” mindset. Prior research has investigated the relationship between growth mindsets and self-efficacy in other contexts. For example, Burnette and colleagues (2019) studied the relationship between growth mindset and entrepreneurial self-efficacy for students. In the study, some students were assigned to view modules across multiple weeks that reinforced the idea of a growth mindset, in which entrepreneurial ability could increase over time. Other students were assigned to an attention-matched control condition where they learned about entrepreneurship myths. Groups did not differ on entrepreneurial self-efficacy or growth mindset at the time of the pretest, but individuals assigned to the growth mindset condition reported significantly greater growth mindset and entrepreneurial self-efficacy scores at post-test and later follow-up than those in the control condition.

While these results are primarily a reflection of the effects of a growth mindset *intervention*, and not of growth mindset itself, on entrepreneurial self-efficacy, they do provide useful considerations. In particular, future work could examine whether increasing VRLSE is more effective for individuals with greater growth mindsets. This investigation would better inform decisions surround VRLSE interventions and whether the growth mindset of individuals should also be bolstered to indirectly support positive training outcomes.

In conclusion, this study offers considerations for both research and practice concerning virtual reality training. As described in this work, the VRLSE scale established by Saville and colleagues (2022) serves as a promising tool for assessing VR training readiness. By deploying this scale and capturing Learning, Interaction, and overall VRLSE scores, organizational leaders can gain a better understanding about learners' confidence to succeed in VR training. This insight may help to predict training success and to identify intervention opportunities aimed at strengthening individuals' confidence to learn using VR. As such, the VRLSE scale has the potential to facilitate successful VR training programs if adopted in organizations.

Table 1.

Summary of VR Training Vignette Conditions

Condition	Skill Type	Learning Outcome
1	Not Specified	Not Specified
2	Technical	Cognitive
3	Technical	Skill-Based
4	Technical	Affective
5	Interpersonal	Cognitive
6	Interpersonal	Skill-Based
7	Interpersonal	Affective

Note. The full text of each vignette may be found in Appendix A.

Table 2.

VRLSE Confirmatory Factor Analysis Model Fit

Models	χ^2	*df*	*p*	CFI	TLI	RMSEA	SRMR
1. Forced Two-Factor	61.50	26	<.01	.97	.95	.11 [.07, .14]	.03
2. Forced One-Factor	94.11	27	<.01	.93	.91	.14 [.11, .18]	.04
3. Hierarchical	61.50	25	<.01	.96	.95	.11 [.08, .15]	.03

Note. Bracketed RMSEA values correspond to a 90% confidence interval at $p = .05$.

Table 3.

VRLSE Model Measurement Equivalence/Invariance Testing

	Model Fit							Comparison						
Models	χ^2	*df*	*p*	CFI	TLI	RMSEA	SRMR	Model	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	ΔCFI
1. Configural Equivalence	412.71	175	<.01	.96	.95	.11[.09, .12]	.03	--	--	--	--	--	--	--
2. Metric Equivalence (first-order)	454.73	217	<.01	.96	.96	.10[.08, .11]	.06	1	454.73	217	.47	42.01	42	0
3. Metric Equivalence (second-order)	454.73	223	<.01	.96	.96	.09[.08, .11]	.06	2	454.73	223	1	.00	6	<.01
4. Scalar Equivalence (first-order)	495.80	265	<.01	.96	.97	.09[.07, .10]	.06	3	495.80	265	.51	41.07	42	0
5. Scalar Equivalence (second-order)	506.07	277	<.01	.96	.97	.08[.07, .10]	.07	4	506.07	277	.59	10.27	12	0

Note. Bracketed RMSEA values correspond to a 90% confidence interval at p = .05.

Table 4.

Descriptive Statistics Across All Conditions

Variable	M(SD)	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1. Gender	--	--							
2. Age	39.47(14.33)	-.06	--						
3. VR Exposure	2.16(.94)	.19**	-.24**	--					
4. First Aid Exposure	2.53(.85)	-.14**	.11**	.07	--				
5. VR Sickness	1.35(.68)	-.05	-.12**	.12**	-.06	--			
6. VRLSE Overall	3.51(.84)	.22**	-.11**	.54**	.04	-.09*	--		
7. VRLSE Learning	3.39(.86)	.16**	-.06	.48**	.08*	-.14*	.96**	--	
8. VRLSE Interaction	3.65(.91)	.25**	-.15**	.56**	.00	-.02	.95**	.82**	--

Note. n = 819. Calculations of average VRLSE scores include participants in all seven conditions. Gender was coded as 0 for female and 1 for male. Due to the low sample size of other gender responses (n = 12), only those who identified as female or male are included in this table.

*p < .05. **p <.01.

Table 5.

Virtual Reality Learning Self-Efficacy Scores by Condition

		Overall Score	Learning Score	Interaction Score
Condition	*n*	M(SD)	M(SD)	M(SD)
Generic VR Training Vignette	120	3.44(.89)	3.31(.91)	3.60(.95)
Interpersonal First Aid / Cognitive	119	3.49(.77)	3.38(.79)	3.61(.86)
Interpersonal First Aid / Skill-based	120	3.52(.80)	3.39(.85)	3.70(.87)
Interpersonal First Aid / Affective	119	3.61(.79)	3.49(.82)	3.76(.84)
Technical First Aid / Cognitive	118	3.57(.76)	3.45(.76)	3.72(.85)
Technical First Aid / Skill-based	119	3.50(.95)	3.43(.95)	3.59(1.02)
Technical First Aid / Affective	116	3.43(.93)	3.29(.96)	3.60(.99)

Note. Scores are calculated as the average of VRLSE items, with a possible range of 1 to 5. Higher scores indicate greater VRLSE.

Table 6.

Virtual Reality Learning Self-Efficacy (Overall) by Skill Type and Learning Outcome Controlling for VR Exposure

Source	SS	*df*	MS	η*2*	*F*	*p*
Corrected Model	179.20	7	25.60	.30	51.03	<.01
Intercept	785.53	1	785.53	.66	1565.95	<.01
VR Exposure	176.13	1	176.13	.30	351.11	<.01
Skill Type	.04	1	.04	<.01	.08	.77
Learning Outcome	.07	2	.03	<.01	.07	.94
Skill Type x Learning Outcome	2.44	2	1.22	.01	2.43	.09
Error	412.84	823	.50			
Total	10820.46	831				
Corrected Total	592.04	830				

Note. R^2 = .30; N = 831.

Table 7.

Virtual Reality Learning Self-Efficacy (Learning) by Skill Type and Learning Outcome Controlling for VR Exposure

Source	SS	*df*	MS	η*2*	*F*	*p*
Corrected Model	149.25	7	21.32	.24	37.37	<.01
Intercept	770.93	1	770.93	.62	1351.07	<.01
VR Exposure	145.72	1	145.72	.24	255.38	<.01
Skill Type	.01	1	.01	<.01	.02	.88
Learning Outcome	.08	2	.04	<.01	.07	.93
Skill Type x Learning Outcome	2.90	2	1.45	.01	2.55	.08
Error	469.61	823	.57			
Total	10177.68	831				
Corrected Total	618.86	830				

Note. R^2 = .24; *N* = 831.

Table 8

Virtual Reality Learning Self-Efficacy (Interaction) by Skill Type and Learning Outcome Controlling for VR Exposure

Source	SS	*df*	MS	η2	*F*	*p*
Corrected Model	221.67	7	31.67	.32	55.48	<.01
Intercept	803.98	1	803.98	.63	1408.64	<.01
VR Exposure	218.19	1	218.19	.32	382.29	<.01
Skill Type	.10	1	.10	<.01	.18	.68
Learning Outcome	.24	2	.12	<.01	.21	.81
Skill Type x Learning Outcome	2.54	2	1.27	.01	2.23	.11
Error	469.73	823	.57			
Total	11788.69	831				
Corrected Total	691.39	830				

Note. R^2 = .32; N = 831.

REFERENCES

Abich, J., Parker, J., Murphy, J. S., & Eudy, M. (2021). A review of the evidence for training effectiveness with virtual reality technology. *Virtual Reality, 25*(4), 919–933. https://doi.org/10.1007/s10055-020-00498-8

Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology, 60*, 451–474.

Alsop, T. (2022, March). *Virtual reality (VR) - statistics & facts*. Statistica. https://www.statista.com/topics/2532/virtual-reality-vr/

American Heart Association (2017, April). AHA mediagenic survey results. Safety and Health Magazine. https://www.safetyandhealthmagazine.com/ext/resources/files/news/Results---Employee-Survey---FINAL.pdf

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., & Calamaras, M. R. (2013). Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 81*(5), 751–760. https://doi.org/10.1037/a0033559

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety, 22*(3), 156–158. https://doi.org/10.1002/da.20090

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W.H. Freeman and Company.

Bandura, A. (1998). Self-efficacy. In H. Friedman (Ed.), *Encyclopedia of mental health.* Academic Press. (Reprinted from V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71–81). Academic Press, 1994.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In *Self-efficacy beliefs of adolescents,* (Vol. 5, Ed. 1, pp. 307–337). Information Age Publishing.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggressions through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, *63*(3), 575–582. https://doi.org/10.1037/h0045925

Bandura, A., & Walters, R. H. (1977). *Social Learning Theory* (Vol. 1). Prentice-Hall.

Barsom, E.Z., Graafland, M. & Schijven, M.P. (2016). Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical Endoscopy*, *30*, 4174–4183 https://doi.org/10.1007/s00464-016-4800-6

Bedwell, W. L., Fiore, S. M., & Salas, E. (2014). Developing the future workforce: An approach for integrating interpersonal skills into the MBA classroom. *Academy of Management Learning & Education, 13*(2), 171–186. https://doi.org/10.5465/amle.2011.0138

Bell, B. S., Tannenbaum, S. I., Ford, J. K., Noe, R. A., & Kraiger, K. (2017). 100 years of training and development research: What we know and where we should go. *Journal of Applied Psychology, 102*(3), 305–323. https://doi.org/10.1037/apl0000142

Bennett-Levy, J., McManus, F., Westling, B. E., & Fennell, M. (2009). Acquiring and refining CBT skills and competencies: Which training methods are perceived to be most effective? *Behavioural and Cognitive Psychotherapy, 37*(5), 571–583. https://doi.org/10.1017/S1352465809990270

Biocca, F., & Delaney, B. (1995). Immersive virtual reality technology. In F. Biocca & B. Delaney (Eds.), *Communication in the age of virtual reality*. Lawrence Erlbaum Associates, Inc.

Bloom, B. S., & Committee of College and University Examiners. (1964). *Taxonomy of educational objectives: The classification of educational goals.* Longmans, Green.

Bouchrika, I. (2021, June 18). *68 training industry statistics: 2021/2022 data, trends & predictions*. Research.com. https://research.com/careers/training-industry-statistics

Burnette, J. L., Pollack, J. M., Forsyth, R. B., Hoyt, C. L., Babij, A. D., Thomas, F. N., & Coy, A. E. (2020). A growth mindset intervention: Enhancing students' entrepreneurial self-efficacy and career development. *Entrepreneurship Theory and Practice*, *44*(5), 878-908. https://doi.org/10.1177/1042258719864293

Busch, T. (1995). Gender differences in self-efficacy and attitudes toward computers. *Journal of Educational Computing Research*, *12*(2), 147–158. https://doi.org/10.2190/H7E1-XMM7-GU9B-3HWR

Byrd, A. P., Stone, R. T., Anderson, R. G., & Woltjer, K. (2015). The use of virtual welding simulators to evaluate experienced welders. *Welding Journal*, *94*(12), 389–395.

Casper, W., Edwards, B. D., Wallace, J. C., Landis, R. S., & Fife, D. A. (2020). Selecting response anchors with equal intervals for summated rating scales. *Journal of Applied Psychology*, *105*(4), 390–409. https://doi.org/10.1037/apl0000444

Chen, I. S. (2017). Computer self-efficacy, learning performance, and the mediating role of learning engagement. *Computers in Human Behavior, 72*, 362–370. https://doi.org/10.1016/j.chb.2017.02.059

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Collins, D. B., & Holton III, E. F. (2004). The effectiveness of managerial leadership development programs: A meta-analysis of studies from 1982 to 2001. *Human Resource Development Quarterly, 15*(2), 217–248. https://doi.org/10.1002/hrdq.1099

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), 678–707. https://doi.org/10.1037/0021-9010.85.5.678

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 189–211. https://doi.org/10.2307/249688

Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite sampling properties of the point estimates and confidence intervals of the RMSEA. *Sociological Methods & Research, 32*(2), 208-252. https://doi.org/10.1177/0049124103256130

Farrell, K. J., Weathers, K. C., Sparks, S. H., Brentrup, J. A., Carey, C. C., Dietze, M. C., Foster, J. R., Grayson, K. L., Matthes, J. C., & SanClements, M. D. (2021). Training macrosystems scientists requires both interpersonal and technical skills. *Frontiers in Ecology and the Environment, 19*(1), 39–46. https://doi.org/10.1002/fee.2287

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Fletcher, G. C. L., McGeorge, P., Flin, R. H., Glavin, R. J., & Maran, N. J. (2002). The role of non-technical skills in anaesthesia: a review of current literature. *British Journal of Anaesthesia, 88*(3), 418–429. https://doi.org/10.1093/bja/88.3.418

Gagne, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist, 39*(4), 377–385. https://doi.org/10.1037/0003-066X.39.4.377

Gist, M. E. (1987). Self-efficacy: Implications for organizational behavior and human resource management. *Academy of Management Review, 12*, 472–485. https://doi.org/10.5465/amr.1987.4306562

Gist, M. E. (1989). The influence of training method on self-efficacy and idea generation among managers. *Personnel Psychology, 42*(4), 787–805. https://doi.org/10.1111/j.1744-6570.1989.tb00675.x.

Gist, M. E., Schwoerer, C., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology, 74*(6), 884–891. https://doi.org/10.1037/0021-9010.74.6.884

Gist, M. E., Stevens, C. K., & Bavetta, A. G. (1991). Effects of self-efficacy and post-training intervention on the acquisition and maintenance of complex interpersonal skills. *Personnel Psychology, 44*, 837–861. https://doi.org/10.1111/j.1744-6570.1991.tb00701.x.

Goldstein, I. L. & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Wadsworth/Thomson Learning.

Graafland, M., Bemelman, W.A. & Schijven, M.P. (2017). Game-based training improves the surgeon's situational awareness in the operation room: a randomized controlled trial. *Surgical Endoscopy, 31*, 4093–4101 https://doi.org/10.1007/s00464-017-5456-6

Gutiérrez, J., Ferrer-García, M., Plasanjuanelo, J., Andrés-Pueyo, A., & Talarn-Caparrós, A. (2016). Virtual reality to train diagnostic skills in eating disorders. Comparison of two low cost systems. In B. K. Wiederhold, G. Riva, & M. D. Wiederhold (Eds.), *Annual review of cybertherapy and telemedicine 2015: Virtual reality in healthcare: Medical Simulation and Experiential Interface, 219*, (pp. 75–81). IOS Press.

Hendrix, M., Al-Sherbaz, A., & Victoria, B. (2016). Game based cyber security training: are serious games suitable for cyber security training? *International Journal of Serious Games*, *3*(1), 53–61. https://doi.org/10.17083/ijsg.v3i1.107

Hill, T., Smith, N. D., & Mann, M. F. (1987). Role of efficacy expectations in predicting the decision to use advanced technologies: The case of computers. *Journal of Applied Psychology*, *72*(2), 307–313. https://doi.org/10.1037/0021-9010.72.2.307

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121. https://doi.org/10.1177/109442819800100106

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huffman, A. H., Whetten, J., & Huffman, W. H. (2013). Using technology in higher education: The influence of gender roles on technology self-efficacy. *Computers in Human Behavior*, *29*(4), 1779–1786.

Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research, 21*(3), 138–151.

Hwang, Y., Lee, Y., & Shin, D.-H. (2016). The role of goal awareness and information technology self-efficacy on job satisfaction of healthcare system users. *Behaviour & Information Technology*, *35*(7), 548–558. https://doi.org/10.1080/0144929X.2016.1171396

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*, 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory - Versions 4a and 54.* University of California, Berkeley, Institute of Personality and Social Research.

Jokisch, M. R., Schmidt, L. I., Doh, M., Marquard, M., & Wahl, H.-W. (2020). The role of internet self-efficacy, innovativeness and technology avoidance in breadth of internet use: Comparing older technology experts and non-experts. *Computers in Human Behavior, 111*. https://doi.org/10.1016/j.chb.2020.106408

Judge, T. A. (2009). Core self-evaluations and work success. *Current Directions in Psychological Science, 18*(1), 58-62. https://doi.org/10.1111/j.1467-8721.2009.01606.x

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of applied Psychology, 86*(1), 80. https://doi.org/10.1037//0021-9010.86.1.80

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors, 63*(4), 706–726. https://doi.org/10.1177/0018720820904229

Kass, K. D. (2014). *Computer self-efficacy: Instructor and student perspectives in a university setting* [Doctoral dissertation, Iowa State University] (pp. 14183–14303). Graduate Theses and Dissertations. https://doi.org/10.31274/ETD-180810-3731

Kirkpatrick, D. L. (1967). Evaluation of training. In R. L. Craig, & L. R. Bittel (Eds.), *Training and development handbook* (pp. 87–112). McGraw-Hill.

Ku, J., Han, K., Lee, H. R., Jang, H. J., Kim, K. U., Park, S. H., Kim, J. J., Kim, C. H., Kim, Y., & Kim, S. I. (2007). VR-based conversation training program for patients with schizophrenia: a preliminary clinical trial. *Cyberpsychology & Behavior*, *10*(4), 567–574. https://doi.org/10.1089/cpb.2007.9989

Lackey, S. J., Salcedo, J. N., Szalma, J. L., & Hancock, P. A. (2016). The stress and workload of virtual reality training: the effects of presence, immersion and flow. *Ergonomics*, *59*(8), 1060–1072. https://doi.org/10.1080/00140139.2015.1122234

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research, 51*(2-3), 220-239. https://doi.org/10.1080/00273171.2015.1134306

Latham, G. P. (1989). Behavioral approaches to the training and learning process. In I. L. Goldstein (Ed.), *Training and development in organizations* (pp. 256–295). Jossey-Bass.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*(2), 460–468. https://doi.org/10.1037/a0025741

Li, L., Yu, F., Shi, D., Shi, J., Tian, Z., Yang, J., Wang, X., & Jiang, Q. (2017). Application of virtual reality technology in clinical medicine. *American Journal of Translational Research, 9*(9), 3867–3880. PMCID: PMC5622235

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84. https://doi.org/10.1037/1082-989X.4.1.84

Matturro, G., Raschetti, F., & Fontán, C. (2019). A systematic mapping study on soft skills in software engineering. *Journal of Universal Computer Science, 25*(1), 16–41. https://doi.org/10.3217/jucs-025-01-0016

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–456. https://doi.org/10.1037/a0028085

Menon, D., Chandrasekhar, M., Kosztin, D., & Steinhoff, D. (2017). Examining preservice elementary teachers' technology self-efficacy: Impact of mobile technology-based physics curriculum. *Contemporary Issues in Technology and Teacher Education, 17*(3), 336–359. https://eric.ed.gov/?id=EJ1154206

Miltiadou, M., & Yu, C. H. (2000). *Validation of the online technologies self-efficacy scale (OTSES)* (ED445672). ERIC. https://eric.ed.gov/?id=ED445672

Moorthy, K., Munz, Y., Sarker, S. K., & Darzi, A. (2003). Objective assessment of technical skills in surgery. *BMJ, 327*(7422), 1032–1037. https://doi.org/10.1136/bmj.327.7422.1032

Noe, R. A. (1986). Trainees' attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review, 11*(4), 736–749. https://doi.org/10.5465/amr.1986.4283922

Noe, R. A., Clarke, A. D. M., & Klein, H. J. (2014). Learning in the twenty-first-century workplace. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 245–275. https://dx.doi.org/10.1146/annurev-orgpsych-031413-091321

Nunnally, J. C. (1978). An overview of psychological measurement. In B. B. Wolman (Ed.). *Clinical diagnosis of mental disorders* (pp. 97–146). Springer. https://doi.org/10.1007/978-1-4684-2490-4_4

Osti, F., de Amicis, R., Sanchez, C. A., Tilt, A. B., Prather, E., & Liverani, A. (2021). A VR training system for learning and skills development for construction workers. *Virtual Reality, 25*(2), 523-–538.

Prolific, Inc. (2023). *Prolific*. https://www.prolific.co

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Qualtrics. (2023) *Qualtrics*. https://www.qualtrics.com

Quinones, M. A. (1995). Pretraining context effects: Training assignment as feedback. *Journal of Applied Psychology, 80*(2), 226–238. https://doi.org/10.1037/0021-9010.80.2.226

Quiñones M. A. (1997). Contextual influences on training effectiveness. In M.A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace* (pp. 177–200). American Psychological Association.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers and Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Rasheed, F., Onkar, P., & Narula, M. (2015). Immersive virtual reality to enhance the spatial awareness of students. In *Proceedings of the 7th international conference on HCI, India HCI 2015* (pp. 154–160). Association for Computing Machinery. https://doi.org/10.1145/2835966.2836288

Ropelato, S., Zünd, F., Magnenat, S., Menozzi, M., & Sumner, R. (2018). Adaptive tutoring on a virtual reality driving simulator. *International SERIES on Information Systems and Management in Creative eMedia (CreMedia), 2017*(2), 12–17. https://doi.org/10.3929/ethz-b-000195951

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest*, *13*(2), 74–101. https://doi.org/10.1177/1529100612436661

Saville, J. D., & Foster, L. L. (2021). Does technology self-efficacy influence the effect of training presentation mode on training self-efficacy? *Computers in Human Behavior Reports, 4*, 100124. https://doi.org/10.1016/j.chbr.2021.100124

Saville, J. D., Noble, S. M., & Foster, L. L. (2022). *Validating the Virtual Reality Learning Self-Efficacy Scale*. Unpublished manuscript.

Schmid Mast, M., Kleinlogel, E. P., Tur, B., & Bachmann, M. (2018). The future of interpersonal skills development: Immersive virtual reality training with virtual humans. *Human Resource Development Quarterly, 29*(2), 125–141. https://doi.org/10.1002/hrdq.21307

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210–222. https://doi.org/10.1016/j.hrmr.2008.03.003

Schunk, D. H. (1987). Self-efficacy and motivated learning. In N. Hastings & J. Schwieso (Eds.). *New directions in educational psychology* (2nd ed.) (pp. 233–251). Falmer Press.

Schunk, D. H. (1995). Self-efficacy and education and instruction. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 281–303). Springer. https://doi.org/10.1007/978-1-4419-6868-5_10

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement, 79*(2), 310-334. https://doi.org/10.1177/0013164418783530

Sousa, M. J., & Rocha, Á. (2019). Leadership styles and skills developed through game-based learning. *Journal of Business Research, 94*, 360–366. https://doi.org/10.1016/j.jbusres.2018.01.057

Spain, R., Penilla, C., Ozer, E., Taylor, R., Ringstaff, C., & Lester, J. (2021). Leveraging game-based learning technologies to introduce adolescents to health science careers during the COVID-19 pandemic. *Journal of STEM Outreach, 4*(4), 1–11. https://doi.org/10.15695/jstem/v4i4.09

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology, 67*, 415-437. https://doi.org/10.1146/annurev-psych-073115-103235

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. https://doi.org/10.1037/0033-2909.124.2.240

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797. https://doi.org/10.1037/0022-3514.69.5.797

Steinmetz, H. (2013). Analyzing observed composite differences across groups. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *9*(1), 1–12. https://doi.org/10.1027/1614-2241/a000049

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences*, *61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfillment on the development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology*, *76*(6), 759–769. https://doi.org/10.1037/0021-9010.76.6.759

Training Magazine (2021, November). *2021 Training Industry Report.* L. Freifield (Ed.). https://trainingmag.com/2021-training-industry-report/

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, *27*(3), 451–481.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, *57*(2), 307–333. https://doi.org/10.2307/1912557

Wang, C. H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education, 34*(3), 302–323. https://doi.org/10.1080/01587919.2013.835779.

Wang, M., Callaghan, V., Bernhardt, J. et al. (2018). Augmented reality in education and training: pedagogical approaches and illustrative case studies. *Journal of Ambient Intelligence and Humanized Computing, 9*, 1391–1402 https://doi.org/10.1007/s12652-017-0547-8

Wang, S., Parsons, M., Stone-McLean, J., Rogers, P., Boyd, S., Hoover, K., Meruvia-Pastor, O., Gong, M., & Smith, A. (2017). Augmented reality as a telemedicine platform for remote procedural training. *Sensors, 17*(10), 2294. https://doi.org/10.3390/s17102294

Whitley Jr, B. E. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. *Computers in Human Behavior, 13*(1), 1–22.

Wilfong, J. D. (2006). Computer anxiety and anger: The impact of computer use, computer experience, and self-efficacy beliefs. *Computers in Human Behavior, 22*(6), 1001–1011. https://doi.org/10.1016/j.chb.2004.03.020

Young, D. (1996). The relative importance of technical and interpersonal skills for new information systems personnel. *Journal of Computer Information Systems, 36*(4), 66–71. https://doi.org/10.1080/08874417.1996.11647229

Yule, S., Flin, R., Paterson-Brown, S., & Maran, N. (2006). Non-technical skills for surgeons in the operating room: a review of the literature. *Surgery, 139*(2), 140–149. https://doi.org/10.1016/j.surg.2005.06.017

APPENDICES

Appendix A

Virtual Reality Learning Self-Efficacy Vignettes

This appendix includes the seven VR training vignettes for Virtual Reality Learning Self-Efficacy (Saville et al., 2022). Each participant was randomly assigned to read and respond to one vignette. One vignette was generic and did not specify a content area, skill type, or learning outcome. The remaining six vignettes focused on the domain of first aid and include one skill type (interpersonal, technical) and one learning outcome (cognitive, skill-based, affective). Each vignette was followed by equivalent instructions for providing VRLSE item ratings.

Generic VR Training Vignette

Imagine you are a participant in a training program. The training program uses immersive Virtual Reality (VR) to teach participants new things. Training using immersive VR has each trainee put on a headset directly over their eyes and use two handheld controllers to interact with a simulated environment.

Picture yourself in the training program just described. For the following items, please rate your confidence on a scale of "Not at all confident" to "Totally confident."

How confident are you that you could…

Technical First Aid, Cognitive

Expertise in first aid requires different knowledge, skills, and attitudes. Some are technical and hands-on, while others are interpersonal, necessitating effective interactions with other people. Imagine you are a participant in a training program to learn information about the **technical** aspects of first aid. The training program uses immersive Virtual Reality (VR) for learning. Training using immersive VR has each learner put on a headset directly over their eyes and use two handheld controllers to interact with a simulated environment. The training will cover information such as the name and location of different parts of the body that are affected when someone is choking. Learning will be assessed after the training via a multiple-choice knowledge test where you are required to answer questions about terminology, principles, and theories underlying the **technical** aspects of first aid.

Picture yourself in the training program just described. For the following items, please rate your confidence on a scale of “Not at all confident" to "Totally confident."

How confident are you that you could…

Technical First Aid, Skill-based

Expertise in first aid requires different knowledge, skills, and attitudes. Some are technical and hands-on, while others are interpersonal, necessitating effective interactions with other people. Imagine you are a participant in a training program to learn **technical**, hands-on first aid skills. The training program uses immersive Virtual Reality (VR) for learning. Training using immersive VR has each learner put on a headset directly over their eyes and use two handheld controllers to interact with a simulated environment. The training will cover technical skills such as strategies to assist someone who is choking, depending on characteristics of the situation. Learning will be assessed after the training via a role-play scenario where you are required to demonstrate **technical** skills taught in the training program.

Picture yourself in the training program just described. For the following items, please rate your confidence on a scale of "Not at all confident" to "Totally confident."

How confident are you that you could…

Technical First Aid, Affective

Expertise in first aid requires different knowledge, skills, and attitudes. Some are technical and hands-on, while others are interpersonal, necessitating effective interactions with other people. Imagine you are a participant in a training program focused on attitudes related to the **technical** aspects of first aid. The training program uses immersive Virtual Reality (VR) for learning. Training using immersive VR has each learner put on a headset directly over their eyes and use two handheld controllers to interact with a simulated environment. The training seeks to boost confidence for doing things such as applying any known technical first aid techniques when someone is choking. Learning will be assessed after the training via a survey asking you to rate your attitudes related to engaging in **technical** aspects of first aid.

Picture yourself in the training program just described. For the following items, please rate your confidence on a scale of "Not at all confident" to "Totally confident."

How confident are you that you could…

Interpersonal First Aid, Cognitive

Expertise in first aid requires different knowledge, skills, and attitudes. Some are technical and hands-on, while others are interpersonal, necessitating effective interactions with other people. Imagine you are a participant in a training program to learn information about the **interpersonal** aspects of first aid. The training program uses immersive Virtual Reality (VR) for learning. Training using immersive VR has each learner put on a headset directly over their eyes and use two handheld controllers to interact with a simulated environment. The training will cover information such as why it is important to obtain consent for helping a person in distress, as well as the laws surrounding the consent process. Learning will be assessed after the training via a multiple-choice knowledge test where you are required to answer questions about terminology, principles, and theories underlying the **interpersonal** aspects of first aid.

Picture yourself in the training program just described. For the following items, please rate your confidence on a scale of "Not at all confident" to "Totally confident."

How confident are you that you could…

Interpersonal First Aid, Skill-based

Expertise in first aid requires different knowledge, skills, and attitudes. Some are technical and hands-on, while others are interpersonal, necessitating effective interactions with other people. Imagine you are a participant in a training program to learn **interpersonal** first aid skills. The training program uses immersive Virtual Reality (VR) for learning. Training using immersive VR has each learner put on a headset directly over their eyes and use two handheld controllers to interact with a simulated environment. The training will cover interpersonal skills such as how to obtain consent to provide care to someone in distress, including what to say and how to say it under different circumstances. Learning will be assessed after the training via a role-play scenario where you are required to demonstrate **interpersonal** skills taught in the training program.

Picture yourself in the training program just described. For the following items, please rate your confidence on a scale of "Not at all confident" to "Totally confident."

How confident are you that you could…

Interpersonal First Aid, Affective

Expertise in first aid requires different knowledge, skills, and attitudes. Some are technical and hands-on, while others are interpersonal, necessitating effective interactions with other people. Imagine you are a participant in a training program focused on attitudes related to the **interpersonal** aspects of first aid. The training program uses immersive Virtual Reality (VR) for learning. Training using immersive VR has each learner put on a headset directly over their eyes and use two handheld controllers to interact with a simulated environment. The training seeks to boost confidence for doing things such as approaching someone in distress to ask for their consent to provide care. Learning will be assessed after the training via a survey asking you to rate your attitudes related to engaging in **interpersonal** aspects of first aid.

Picture yourself in the training program just described. For the following items, please rate your confidence on a scale of "Not at all confident" to "Totally confident."

How confident are you that you could…

Appendix B

Survey Items

This appendix includes all survey items excluding VRLSE items, which may be found in Appendix C.

Participant Exclusion Criteria

This item was asked alongside informed consent:

1. Are you a student or employee of [UNIVERSITY]?
   a. No
   b. Yes

This item was randomly inserted among VRLSE scale items:

2. Please select "A little confident"
   a. Not at all confident
   b. A little confident
   c. Somewhat confident
   d. Very confident
   e. Totally confident

This item was asked following all other survey items:

3. Please consider how carefully you have responded to items in this study. In your honest opinion, should we use your data? You will be compensated regardless of your answer, but your honesty helps to improve our work.
   a. No
   b. Yes

Manipulation Checks

These items were shown immediately after the initial vignette was presented. This allowed participants to refer to the vignette while answering. Instructions preceding the manipulation check items are as follows:

"Please answer these two questions based on the training scenario above. The correct answer for each item will be required to ensure that you understand specific elements about the training."

1. What topic will the training cover?
    a. No topic specified
    b. Technical aspects of first-aid
    c. Interpersonal aspects of first-aid
2. How will learning be assessed in the training?
    a. No learning assessment specified
    b. Multiple-choice knowledge test about first-aid terminology, principles, and theories
    c. Role-play scenario demonstrating skills taught in the training
    d. Survey asking for ratings about attitudes

Demographic Items

1. Which gender do you most closely identify with?
    a. Female
    b. Male
    c. Non-binary
    d. Self-describe: _____

2. What is your ethnicity?
    a. Not Hispanic or Latino/a/x
    b. Hispanic or Latino/a/x
3. Which race do you most closely identify with?
    a. African, African American, or Black
    b. Asian or Asian American
    c. European, European American, or White
    d. Middle Eastern, North African, or Arab American
    e. Native American or Alaskan Native
    f. Native Hawaiian or Pacific Islander
    g. Multi-racial
    h. Self-describe: ________
4. What is your age?

This item asked participants to select their age from a drop-down list. Response options include "younger than 18," each numeric year from 18 to 100, and "100+."

5. What is the highest level of education you have attained ***so far***?
    a. Some high school
    b. High school diploma or equivalent
    c. Vocational training
    d. Some college
    e. Associate degree (e.g., AA, AE, AFA, AS, ASN)
    f. Bachelor's degree (e.g., BA, BBA BFA, BS)
    g. Some graduate school

h. Master’s degree (e.g., MA, MBA, MFA, MS, MSW)

i. Applied or professional doctorate degree (e.g., MD, DDC, DDS, JD, PharmD)

j. Doctorate degree (e.g., EdD, PhD)

k. Other (please specify): ________

6. Are you currently a student?

   a. No

   b. Yes, part-time

   c. Yes, full-time

7. What is your current employment status?

   a. Not employed and not currently looking for work

   b. Not employed and currently looking for work

   c. Employed part-time (less than 35 hours per week)

   d. Employed full-time (35 or more hours per week)

   e. Retired

Exposure to Virtual Reality

1. How would you rate your current familiarity with VR?

   a. Not at all familiar

   b. Slightly familiar

   c. Somewhat familiar

   d. Moderately familiar

   e. Extremely familiar

2. How often do you currently use VR?

   a. Never

b. Seldom
c. Some of the time
d. Frequently
e. Constantly

3. How would you rate your current level of expertise in VR?
    a. Very low
    b. Below average
    c. Average
    d. Above average
    e. Very high

VR Experience

1. How many total hours of experience do you have with using VR? ________

    Numeric response: if the response was greater than 0, items 2 - 4 were asked.

2. Have you taken a training that used VR to teach you new things?
    a. No
    b. Yes
3. How many years has it been since you have used VR?

    Dropdown from "less than 1 year" to 99+

4. Have you ever gotten physically sick while using VR?
    a. No
    b. A little bit sick
    c. Somewhat sick
    d. Very sick

e. Extremely sick

This item was open-ended:

5. In your own words, how would you describe your prior exposure or experience with immersive virtual reality (with a headset and controllers)?

First Aid Exposure

1. How would you rate your current familiarity with first aid?
    a. Not at all familiar
    b. Slightly familiar
    c. Somewhat familiar
    d. Moderately familiar
    e. Extremely familiar
2. How often do you currently use first aid?
    a. Never
    b. Seldom
    c. Some of the time
    d. Frequently
    e. Constantly
3. How would you rate your current level of expertise in first aid?
    a. Very low
    b. Below average
    c. Average
    d. Above average
    e. Very high

First Aid Experience

1. Have you had any first aid training before?
    a. No
    b. Yes

If the response to item 1 was yes, item 2 was asked.

2. How many years has it been since you have taken a first aid training?

Dropdown from "less than 1 year" to "99+"

Appendix C

Virtual Reality Learning Self-Efficacy Scale (VRLSE)

For the following items, please rate your confidence on a scale of "Not at all confident" to "Totally confident."

How confident are you that you could…

1. ... without any doubt, effectively use VR to learn about the program topic.
2. ... use VR to learn and master basic concepts related to the program topic on your own.
3. ... use VR to learn successfully.
4. ... complete the program material in a timely manner.
5. ... use VR to learn and master difficult concepts without asking for help.
6. ... easily learn how to operate VR controllers.
7. ... easily learn how to use a VR headset.
8. ... successfully navigate in a VR environment on your own.
9. ... successfully interact with virtual objects in VR.

Note. Items should be randomly ordered and administered on a five-point Likert-type scale, with response options of "Not at all confident" (1), "A little confident" (2), "Somewhat confident" (3), "Very confident" (4), and "Totally confident" (5). Items 1-5 correspond to the Learning factor and items 6-9 correspond to the Interaction factor. Scores can be calculated as the average of Learning items, Interaction items, and all nine items for an overall VRLSE score.