Improved Training for Disasters Using 3-D Virtual Reality Simulation

Sharon Farra¹, Elaine Miller², Nathan Timm², and John Schafer²

Abstract
The purpose of this study was to examine the effects of virtual reality simulation (VRS) on learning outcomes and retention of disaster training. The study used a longitudinal experimental design using two groups and repeated measures. A convenience sample of associate degree nursing students enrolled in a disaster course was randomized into two groups; both groups completed web-based modules; the treatment group also completed a virtually simulated disaster experience. Learning was measured using a 20-question multiple-choice knowledge assessment pre/post and at 2 months following training. Results were analyzed using the generalized linear model. Independent and paired t tests were used to examine the between- and within-participant differences. The main effect of the virtual simulation was strongly significant (p < .0001). The VRS effect demonstrated stability over time. In this preliminary examination, VRS is an instructional method that reinforces learning and improves learning retention.

Keywords
virtual reality, education, disaster planning, mass casualty training

¹Miami University Oxford, OH, USA
²University of Cincinnati, OH, USA

Corresponding Author:
Sharon Farra, Miami University Middletown, 205 R Thesken Hall, 4200 N University Blvd., Middletown, OH 45042, USA.
Email: farrasl@muohio.edu
In 2011, there were 302 worldwide natural disasters that claimed more than 29,780 lives, affected nearly 206 million others, and resulted in record economic damages within the United States of 366 billion dollars (Centre for Research on the Epidemiology of Disasters, 2011). Given the prevalence of disasters, there is an urgent need to improve the education of health care workers preparing for disaster response (Chapman & Arbon, 2008; Slepski & Littleton-Kerney, 2010). One challenge to preparedness is a lack of disaster training opportunities. It is expensive, labor-intensive, and difficult to conduct live exercises, but virtual reality simulation (VRS) may offer an accessible and cost-effective alternative to meet training needs (Heinrichs, Youngblood, Harter, Kusumoto, & Dev, 2010). The continued development of technology, web services, and software engineering has expanded the opportunity to develop VRS to practice disaster response (Chen, Rebolledo-Mendez, Liarokapis, de Freitas, & Parker, 2008; Hansen, 2008; Roy, Sticha, Kraus, & Olsen, 2006).

VRS uses 3-D environments and computer interface to allow participants to interact within a virtual environment (Bergeron, 2008). Moreover, VRS is increasingly being used as a method to educate health care workers to respond to disasters. A few current examples of virtual simulation use in disaster training include (a) the Center for Disease Control and Prevention’s (CDC) recent implementation of virtual reality training for Deployment Safety and Resilience (Klomp, Spitalnick, & Reissman, 2011), (b) the Incident Command Training virtual reality training tools based on the U.S. Department of Homeland Security’s Incident Management System (Barrera, 2008), and (c) the “Play 2 Train,” part of the Idaho Bioterrorism Awareness and Preparedness Program (IBAPP), a virtual world that is designed specifically to meet the distinctive training needs of the professionals who will be on the scene in the event of any large-scale disaster (Roberts, 2010).

**VRS in Disaster Training**

Simulation is an educational process that imitates an environment and requires the learner to demonstrate procedural techniques, decision making, and critical thinking (National Council of State Boards of Nursing, 2005). It is a learning strategy in which elements of the real world are incorporated to achieve specific learning or evaluation outcomes (Gaba, 2004). Decker, Sportsman, Puetz, and Billings (2008) describe simulation as including a variety of strategies such as task trainers, computer-based programs, high-fidelity human patient simulators, standardized patients, and virtual reality. VRSSs are quickly becoming a pervasive simulation learning strategy for nurses. Technologic, programming, and software development had led to a
high level of realism, availability, and the affordability of these systems resulting in increased use in nursing programs (Kilmon, Brown, Ghosh, & Mikitiuk, 2010).

The use of simulation in nursing education is valuable because it allows students to repeatedly participate rare and critical events in a safe environment with no direct risk to patients (Decker et al., 2008). In a systematic review of simulation-based learning in nursing education, Cant and Cooper (2010) discovered that all included studies found that simulation was a valid teaching and learning strategy. Findings of six studies demonstrated gains in knowledge, critical thinking ability, and satisfaction or confidence in simulation learners when compared with control groups. Norman (2012) reinforced these findings in another systematic review of simulation literature finding that simulation was helpful in creating an environment that fostered the development of knowledge, skills, confidence, and safety. Yet, both authors identified the need for continuing study of the transfer of these abilities to the clinical setting.

In reviewing current scientific literature related to VRS and disaster training, three areas of study were identified: (a) descriptions of the participant’s experience in the virtual environment, (b) learning results of participation in the simulation, and (c) an exploration of how knowledge construction occurs in the virtual environment.

Heinrichs, Youngblood, Harter, and Dev (2008) and Kizakevich et al. (2007) evaluated participant VRS experiences in disaster training using ordinal response questionnaires. The authors found that a majority felt immersed and reported an increase in confidence following participation in the simulation. Physicians and nurses in a VRS, caring for disaster victims in an emergency department, felt VRS was useful for learning teamwork and clinical skills (Heinrichs et al., 2010). Vincent, Sherstyuk, Burgess, and Connolly (2008) as well as Wilkerson, Avstreich, Gruppen, Beier, and Wooliscroft (2008) reported that physicians and paramedics gave high evaluation scores to VRS disaster training. However, these studies were limited by the use of convenience and small samples sizes. Minimal information is provided regarding reliability and validity of questionnaires or coding of qualitative data from focus groups or interviews.

To evaluate knowledge acquisition following VRS, Bergeron (2008) administered pre/post and 6-week tests to two groups. VRS participants and traditional methods participants were compared, and significantly greater learning retention was demonstrated by the VRS group at 6 weeks ($n = 89$). Knight et al. (2010) found in research performed in the United Kingdom that triage tagging accuracy and triage step accuracy were significantly higher in a VRS training group compared with a group trained using a card.
sort exercise. Study participants attended a Major Incident Management Support course. In this same study, they found no difference in time to triage between the two groups \((n = 91)\). Although most of the studies used pre/post-testing, limited information is available related to the development, validity, and reliability of these tools.

In a study conducted in the Netherlands, Van der Spek, Wouers, and van Osterdendorp (2010) explored how learning was fostered in the VRS in a small pilot study involving paramedics and triage. The authors measured associated word pairs for mental model elicitation in forming conceptual models to evaluate learner mental model structure pre- and post-VRS completion. The study found no change in mental model construction post-simulation, but the study was limited by size, use of convenience sample, and the use of previously triage trained participants.

As indicated by these limited research efforts, VRS reportedly provides a realistic environment for disaster training. In studies, VRS demonstrated equal or improved learning outcomes immediately post-training to traditional teaching methods. These positive outcomes of VRS may be explained by the learning theory of situated cognition.

**Theoretical Framework**

Situated learning theory supports the use of virtual simulations by increasing learner’s knowledge and retention (Bares, Zettlemoyer, & Lester, 1998). This theory is based on concepts of embodiment (cognition is dependent on the sensorimotor brain and body), embeddedness (cognition is fixed in context specific representations), and extension (cognitive systems exist in a physical and social environment; Robbins & Aydede, 2009). The domain of learning is the culture in which acquired knowledge will be used. For example, the scalpel cannot be used appropriately without understanding the community or culture of the operating room. Meaningful learning will only take place if it is embedded in the culture of the situation (Brown, Collins, & Duguid, 1989). Educational technology, through virtual reality and interactive multimedia, is supported as an avenue to bring situated learning into the classroom (Harley, 1993). According to Herrington and Oliver (1995), virtual simulations “provide a powerful acceptable vehicle for the critical characteristics of a traditional apprenticeship” (p. 2).

**Purpose**

This innovative study examined the effectiveness of virtual reality disaster simulation in fostering knowledge acquisition and retention of disaster training in nursing students.
Two research questions were considered:

**Research Question 1**: Are there differences in knowledge of disaster response between participants who receive web-based disaster training compared with those who receive web-based training and a virtual disaster simulation experience?

**Research Question 2**: Are there differences in retention of knowledge of disaster response between participants who receive web-based disaster training compared with those who receive web-based training and a virtual simulation experience?

**Method**

**Design**

This study was a longitudinal experimental design using two groups and repeated measures.

**Participants**

An *a priori* power analysis was conducted using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2009) to determine the required sample size. For the purposes of feasibility, a sample total of 32 (16 in each group) was preferred to achieve a minimum of 80% power assuming a small effect size, autocorrelation of .2, and alpha of .05. Given the potential for attrition from the study, sample size was increased to a total desired sample of 40. All nursing students in their final year of study in either the capstone courses or pediatrics courses were invited to participate in the disaster. Inclusion criteria were that participants be associate degree nursing students at least 18 years of age and be able to participate in VRS. Exclusion criteria were previous extensive disaster training or paramedic certification. Approval was obtained from the Institutional Review Board of the University and Community College. Participation was voluntary, and information regarding participation or nonparticipation was not shared with instructors nor were responses to assessments of knowledge.

**Knowledge Measure**

Competencies for the disaster course were based on the criteria developed by the International Nursing Coalition for Mass Casualty Education (INCMCE; 2003). Items for the postknowledge assessment measured educational competencies for registered nurses responding to mass casualty incidents (INCMCE,
The formative knowledge assessment tool used in the study consisted of 20 criterion-referenced, multiple-choice questions. In addition, the knowledge assessment tool was developed from an existing author-developed exam. Previous administration of the knowledge assessment reliability was measured using Kuder–Richardson coefficient of reliability (KR20) and was found to be an acceptable $r = .72$ (Miller, Linn, & Gronlund, 2009).

**Validity.** The knowledge assessment tool, in its existing form, was assessed for validity by a panel of three disaster experts and three education experts. Disaster experts were asked to evaluate each question to assess content validity. For each question, the reviewer was given the question and the competency it measured. Reviewers were asked to judge the following criteria derived from Miller and colleagues (2009): item congruence, relevance of item to content domain, accuracy of the knowledge assessment item, inaccuracy, and suitability of distracters.

Education experts reviewed the questions using items derived from Burton, Sudweeks, Merrill, and Wood (1991). Item analysis from previous administrations of the knowledge assessment was provided to reviewers, including the item difficulty, discrimination, and distractor analysis.

**Reliability.** After completion of the postknowledge assessment tool by student participants, two methods of reliability were assessed, including a KR20 (Kuder & Richardson, 1937) and calculation of the test–retest reliability coefficient.

**Web-Based Education.** The first component of the educational intervention was the web-based learning modules. There were five sets of modules developed from the learning objectives recommended by the INCMCE (2003). Modules focused on creating a greater understanding of clinical issues pertaining to direct nursing care of victims. Topics included concepts of detection, personal protection, and immediate care for those affected by disaster. All of the modules consisted of textual narration and embedded active learning strategies such as practice questions, labeling activities, and scenarios.

**Virtual Simulation Intervention.** Second Life was created in 2003 and is an online virtual reality world where users interact with a 3-D environment. The users are represented in the world by an avatar through which they interact within a world where there is spatial geography, movement, sound, and the ability to communicate via chat or note cards (Second Life, 2011). Two VRSs were developed in the Second Life virtual reality platform, where participation was achieved through first-person avatar. The avatar was designed for use within the scenarios and was dressed in the school uniform and designed as gender and race neutral. The disaster scenarios were placed in a private area of Second Life with access available only to study participants, investigators, and consultants. The first scenario consisted of a START triage.
exercise including eight victims with varying injuries who were triaged by the participant. The second VRS was a decontamination exercise in which the participant donned appropriate personal protective equipment and decontaminated the patient moving through the Hot, Warm, and Cold zones. Feedback was provided to the learner throughout the exercises via interactive activities, note cards, and messages.

The virtual simulation was developed to reinforce the web-based learning presented in the online modules. The scenarios took approximately 30 min to complete. The simulation content was developed by the study author and reviewed by three disaster experts for authenticity of activity, correctness of modeling processes, and the appropriateness of roles and perspectives. Three education experts assessed the simulations’ ability to support (a) construction of knowledge, (b) effectiveness of coaching and scaffolding, (c) the scenario’s ability to promote reflection and abstraction, and (d) integration of assessment of the learning tasks (Herrington & Oliver, 1995). After the simulation script had been reviewed and revised, the simulation was articulated into the 3-D environment using the Second Life platform (Second Life, 2011). A tutorial was adapted for use with the module to familiarize the participants with the virtual environment.

**Procedure**

Within this study, the type of disaster training was the independent variable; learning and retention were the dependent variables. The participants were randomly assigned to intervention group (web-based teaching method only) or standard care group (web-based teaching with virtual simulation group) for disaster training. Nursing students in the capstone courses or pediatrics courses were invited to participate in the disaster training study via announcements and email. Students who chose not to participate in the research study were directed toward “Traditional” modules. Participants consented to participate in the study by selecting the “Simulation Study” modules and were randomized into two groups using the learning management system’s (LMS) random team generator.

Prior to participating in the modules, both groups of participants completed a preknowledge assessment of 20 multiple-choice questions. The control group completed the web-based disaster training modules, alone and the treatment group completed the web-based disaster modules and a virtual disaster simulation. After training, both groups of participants again completed the same 20-question multiple-choice knowledge assessment. Two months post disaster training, email reminders were sent to all participants to participate in the final administration of the multiple-choice knowledge assessment.
The procedure for accessing was as follows: after completion of web-based modules, the treatment group was prompted to access the VRS. There was a short orientation to the simulation environment and then students participated in a VRS experience reinforcing the concepts of disaster training. Using a computer with Internet capability, a mouse, and a keyboard, learners accessed two virtual disaster scenarios. Each scenario lasted approximately 10 min. The participant was able to assess, triage, and provide first aid to victims of radioactive and explosive events. The tutorials offered summative and cumulative performance feedback during and after each scenario and reinforced the concepts taught in the disaster course. After completing the simulation, participants completed the online postknowledge assessment (20-question multiple-choice test). The 2-month knowledge assessment was open for completion for 7 days. After completion of all three knowledge assessments (20-question multiple-choice test) within the 2-month time frame, all participants had the option of receiving a Starbucks gift card as compensation for their time.

Findings

Sample Characteristics

Participants were recruited from 2nd-year associate degree nursing students enrolled in their final nursing courses at a community college. Students enrolled in the program ranged in age from 18 to 57 years of age. The student population is 73% White, 18% African American, 7% Asian, and 2% Hispanic. The majority of the students were female 91% and 9% are male (M. Miller, personal communication, May 5, 2010). Seventy-five students were invited to participate in the study; of these, 54 (total response rate of 72%) responded that they would participate in the study. Of those agreeing to participate, 47 participants completed the preassessment and postassessment, and 41 participants (13% attrition) completed all three assessments. Within the sample completing the first two assessments, there were 4 men and 43 women. Ages of participants completing all three assessments were as follows: 10 were 18 to 25 years old, 10 were 26 to 33 years old, 13 were 34 to 41 years old, 7 were 42 to 49 years old, and 1 student was 50 and above.

Instrument Characteristics

To assess score consistency or stability of the knowledge assessment, a Pearson’s product–moment correlation coefficient was computed to assess test–retest reliability. The results demonstrated knowledge score consis-
tency and stability with \( r = .73 \) (Leech, Barrett, & Morgan, 2008; Murphy & Davidshofer, 2005). To assess the internal consistency of the items, KR20 was also computed. A value of \( r = .62 \) was obtained, which is acceptable reliability for a criterion-referenced assessment, especially due to the homogeneity of the population of 2nd-year nursing students and the small number of questions (McGahee & Ball, 2009). According to Oermann and Gaberson (2009), a reliability coefficient of .60 to .85 is desirable for nursing assessments.

Item analysis was conducted on the postknowledge assessment participant data. Three items were answered by all participants correctly; these items were identified prior to administration as mastery items (high priority), and it was anticipated they would be answered correctly (Morrison, Nibert, & Flick, 2006). All other items scores were positive; ranges on the point biserial correlation coefficients of .21 to .68 was interpreted as good to excellent (Morrison et al., 2006). Item difficulty ranged from 31% to 100%.

Analysis

Histograms of variables were plotted and examined. The data were found to be nonnormally distributed, so advanced statistical methods were required. To address the research questions and to assess the overall effect of the VRS, the results of the knowledge scores were analyzed by SPSS software using the generalized linear model, implementing generalized estimating equations. Generalized linear models extend linear regression models to accommodate nonnormal response distributions and transformation of linearity (Diggle, Heagerty, Liang, & Zeger, 2002). The Generalized Estimating Equations (GEE) was used to account for clustering observed in the longitudinal data (McCullagh & Nelder, 1989). Model fit was confirmed by comparing the quasilikelihood information criterion (QIC) of the full model with the QIC of an intercept-only model. A critical alpha level of .05 was used to demonstrate statistical significance.

Findings

The omnibus test of the overall model was calculated using a Wald statistic and was highly significant \( p < .0001 \) indicating that there were significant differences between the VRS (treatment) and nonsimulation (control) groups. Six of the 2-month postknowledge scores were missing due to participant attrition. Therefore, missing data were predicted using a regression equation with preand postknowledge assessment data to predict the missing 2-month knowledge scores. No substantive differences were found in the statistical significance of
the model when using the imputed data compared with the model created from the original data set with missing data. Overall, the main effect of the virtual simulation was strongly significant \( p < .0001 \). Although the initial independent \( t \) test of the preknowledge assessment scores demonstrated that the two groups differed significantly, the use of GEE controlled for these differences when the research question was answered. Both groups showed similar rate of improvement of knowledge assessment scores following the teaching intervention in the first postknowledge assessment. The significant differences can be seen in the knowledge assessment scores at 2 months post-intervention (see Figure 1 and Table 1). The virtual simulation effect demonstrated significant stability over time. The nonsimulation (control) group showed significant decay in knowledge scores.

**Within- and Between-Participant Differences**

Individual knowledge scores between participants were compared using independent sample \( t \) tests. Significant differences were found between the two preknowledge scores groups, \( p = .023 \). Although it is considered optimal to have no differences initially in an experimental design, it is not uncommon for these differences to occur (randomization failure). In such cases, it is common practice to statistically control for these differences.
using such methods as GEE, latent growth curve analysis, and multilevel models. Knowledge scores between the immediate postassessment groups differed significantly \((p = .021)\) and group differences in the knowledge scores on the 2-month assessment were significant \((p < .0001)\). The VRS group knowledge scores were higher in each instance. The within-group effects of the two groups were explored using paired \(t\) tests. There were significant differences between the pre- and postknowledge scores \((p < .001)\), the Cohen’s \(d = .359\), the postknowledge scores, and 2-month knowledge scores \((p < .0001)\) with an effect size of \(d = -.676\). Paired samples correlations indicated strong correlation between postknowledge scores and 2-month knowledge scores \((p < .0001)\).

Effect sizes for the analyses with 95% confidence intervals as Cohen’s \(d\) are presented in Table 2. Effect sizes for all of the comparisons between the groups at each time point are large. Within-group comparisons between the time points for the intervention group have the effects of the control group partialled out, and the first comparison has a medium effect, whereas the second is in Cohen’s “large” category. The last within comparison includes 0 in the confidence interval.

**Discussion**

In this preliminary study, the VRS had a strong positive effect on retention of disaster training. When looking at the immediate postassessment and
2-month postknowledge scores, a significant difference was found in the stability of the knowledge scores in the VRS group over time. The VRS group demonstrated higher knowledge scores on the immediate post-assessment. The VRS group demonstrated significantly higher knowledge scores on the 2-month postknowledge assessment and demonstrated improved retention ($p < .0001$).

These results support those of Bergeron (2008) who found knowledge retention scores were stable for the VRS groups with decay in scores for the control group 6 weeks post-training. In addition, participant experiences with the VRS were similar to other studies in that participants reported that VRS is a realistic and positive learning experience (Heinrichs et al., 2008; Heinrichs et al., 2010; Kizakevich et al., 2007; Vincent et al., 2008).

Limitations of the study include use of convenience sampling, sample size, and the length and number of assessments of the study. The initial randomization of the two groups failed resulting in significant differences between the groups at baseline, which were controlled for in the analysis by use of GEE. Only cognitive testing was used to measure learning and retention. Testing within the psychomotor and affective domains is needed. The VRS group supplied course evaluations of the VRS experience. On end of course evaluations, 80% of participants gave positive comments about participation in the VRS. Positive comments included multiple references to applying information in a visual way not available in a passive module setting, realistic experiences, and a better understanding of zones of triage. A few of the participants were motivated enough to email the primary investigator regarding the VRS experiences. Although some were enthusiastic, other participants gave feedback that they had difficulty navigating the VRS. Reasons given included that the controls and operating of teleports within the VRS were difficult to maneuver. Only one of the participants provided negative feedback reporting that the VRS was not realistic.

### Table 2. Effect Size: Between-Group Differences.

<table>
<thead>
<tr>
<th>Knowledge</th>
<th>Experimental Group</th>
<th>Control Group</th>
<th>Cohen's d</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preknowledge scores</td>
<td>22 13.5 13.5</td>
<td>25 11.3 2.021</td>
<td>.964</td>
<td>[.359, 1.569]</td>
</tr>
<tr>
<td>Postknowledge scores</td>
<td>22 17.68 17.68</td>
<td>25 16.24 2.134</td>
<td>.741</td>
<td>[.149, 1.334]</td>
</tr>
<tr>
<td>Two-month knowledge scores</td>
<td>21 16.95 16.95</td>
<td>20 14.1 2.49</td>
<td>1.284</td>
<td>[.612, 1.957]</td>
</tr>
</tbody>
</table>

Note: CI = confidence interval.

*Within-group differences are found in text.*
Farra et al.

During the VRS, participant progression through the scenarios was tracked. The tracking gave more evidence that the scenario was difficult to manipulate. VRS participants did not complete all aspects of the simulation uniformly. Twenty-four percent ($n = 6$) of the VRS participants completed fewer than half of the 15 stations within the VRS. Despite the failure to interact with all the patients in the triage scenarios, the VRS still significantly affected learning retention at 2 months. The reasons for these failures to complete were not assessed by the study design, but it is assumed that the reasons listed above related to navigation within the system are at least part of the cause.

Benefits of VRS include the ability to replicate actual buildings and areas into realistic scenarios. A wide variety of disaster situations can be modeled, including man-made and natural disasters. Participants may repeat the scenarios at any time. Their performance can be tracked and evaluated throughout the scenario. In future studies, the development of the VRS must offer the participant ease of navigation and strong feelings of embeddedness. The VRS in this study was from the focus of a single player view, whereas multiplayer simulations allow for interaction between participants that may strengthen the learning experience by improving extension (interaction between environment and others) within the VRS.

In this preliminary examination, VRS is an instructional method that reinforces learning and improves learning retention. The use of VRS in disaster training may improve accessibility and cost-effectiveness as an alternative, or supplement, to live drills. Further study is necessary in the use of VRS in disaster training and an exploration of areas of learning where VRS may be an applicable pedagogy in nursing education. Larger studies involving more participants and varied VRS delivery methods are essential. Studies are needed with more longitudinal data to explore the stability of the learning effect. All three domains of learning need to be examined when testing the effects of VRS. Evaluations of cognitive, affective, and psychomotor responses to VRS are needed so that researchers can examine the translation knowledge of VRS content to live disaster situations. It is important to be able to demonstrate that cognitive knowledge transforms into the willingness and ability to perform well in disaster situations. Further improvements of VRS may improve the results of this study.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors gratefully acknowledge
Sigma Theta Tau and the METI Corporation for their support of this research through the Sigma Theta Tau METI Simulation, Nursing Research Grant 6330.

Clinical Resources

Centers for Disease Control and Prevention: Disaster Site Management http://www.cdc.gov/niosh/topics/emres/sitemgt.html

Centers for Disease Control and Prevention: Emergency Preparedness and Response http://emergency.cdc.gov


Nursing Emergency Preparedness Education Coalition http://www.nursing.vanderbilt.edu/incmce/overview.html

References


Farra et al.


training: A pragmatic controlled trial. *Resuscitation*, 81, 1175. doi:10.1016/j.resuscitation.2010.03.042


