

Distributed Immersive Virtual Reality Simulation Development for Medical Education

Dale C. Alverson, M.D.,¹ Stanley M. Saiki Jr, M.D.,^{4,8} Thomas P. Caudell, Ph.D.,²
Kenneth Summers, Ph.D.,² Panaiotis, Ph.D.,² Andrei Sherstyuk, Ph.D.,⁴ David
Nickles, M.S.,⁴ James Holten, III,² Timothy E. Goldsmith, Ph.D.,³ Susan M. Stevens,
M.S.,³ Kathleen Kihmm,⁴ Stewart Mennin, Ph.D.,¹ Summers Kalishman, Ph.D.,¹ Jan
Mines, M.A.,¹ Lisa Serna,¹ Steven Mitchell, M.D.,¹ Marlene Lindberg Ph.D.,⁴ Joshua
Jacobs, M.D.,⁴ Curtis Nakatsu, M.D.,⁴ Scott Lozanoff, Ph.D.,⁴ Diane S. Wax, M.P.A.,
M.B.A.,¹ Linda Saland, Ph.D.,¹ Jeffrey Norenberg, PharmD.,⁵ George Shuster,
DNsc.,⁶ Marcus Keep, M.D.,¹ Rex Baker, M.D.,¹ Holly S. Buchanan, Ed.D.,¹ Randall
Stewart, M.D.,¹ Mark Bowyer, M.D.,⁷ Alan Liu, Ph.D.,⁷ Gilbert Muniz, Ph.D.,⁷
Robert Coulter, M.A.,¹ Christina Maris,¹ David Wilks, M.D.¹

¹School of Medicine, ²School of Engineering, ³Department of Psychology, ⁵College of Pharmacy, and
⁶College of Nursing

^{1, 2, 3, 5, 6}University of New Mexico, Albuquerque, New Mexico 87131 U.S.A.

⁴John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii 96822 U.S.A.

⁷National Capital Area Simulation Center, Uniformed Services University of the Health Sciences, Silver
Springs, Maryland 20910 U.S.A.

⁸Pacific Telehealth and Technology Hui, Tripler Army Medical Center, Honolulu, Hawaii 96859 U.S.A.

Phone: (+)1-505-272-8633

Fax: (+)1-505-272-0800

Email: dalverson@salud.unm.edu

ABSTRACT

Training professionals for real-world application of required knowledge and skills and assessing their competence are major challenges. Simulations are being used in education and training to enhance understanding, improve performance, and assess competence. Validated virtual reality (VR) simulations provide a means of making experiential learning reproducible and reusable. Advanced communication networks, such as Internet2 Access Grid, allow dissemination of these simulations and collaborative learning independent of distance. The prior experiences of our three universities led to an interdisciplinary collaboration to further develop and evaluate an integrated, fully immersive, interactive VR based system. This environment employs simulations that are visually three-dimensional and are driven dynamically by a rules-based artificial intelligence engine within Flatland, a virtual environments development software tool, and associated commodity hardware. Studies include usability and validation, deployment for distributed testing over Internet2, and evaluation of impact on training and performance using concept mapping and knowledge structure methods. Subject matter experts found face and content validity in our closed head injury simulation. Seven pairs of medical students participated collaboratively in problem solving and managing of the simulated patient in VR. Students stated that opportunities to make mistakes and repeat actions in VR were extremely helpful in learning specific principles and they felt more engaged than in standard text-based scenarios. 48 students participated in knowledge structure experiments pre and post simulation experiences. Knowledge structure relatedness ratings were significantly improved in those students with lower pre-VR relatedness ratings indicating a potential value of VR simulation in learning. This research cuts across the integration of computing, networking, human-computer interfaces, learning, and knowledge acquisition. VR creates a safe environment to make mistakes and could allow rapid deployment for just-in-time training or performance assessment.

INTRODUCTION

The vast amount of existing and emerging new knowledge in the health related sciences create new challenges in medical education. Furthermore, there are several medical science concepts that are difficult for learners to comprehend and educators to teach.¹ Developing methods to determine adequate acquisition, retention and competence in the application of those concepts and knowledge, as well as attainment of appropriate clinical skills, continues to be a major critical endeavor in medicine² as efforts to decrease medical errors and improve quality of care have reached high levels of public interest.^{3,4}

Simulations have been used as a method to enhance learning, training and assessment of competence. In a detailed analysis of the literature and review of military simulation efforts, Champion and Higgins⁵ concluded that simulation is an effective and cost efficient approach to training military personnel, enhancing knowledge transfer, and improving performance. For example, they reported that flight simulators have been shown to be effective for training and improving subsequent performance and that simulation, if designed and integrated appropriately, could be applied effectively in training combat medics and physicians. Similar conclusions were reported by Satava and Jones⁶ regarding the potential value of using virtual reality to assess competence. Ziv, et. al.,⁷ also make a compelling argument of the need to develop further simulation-based medical education as an ethical imperative to ensure optimal treatment, patient safety and well being. They argue an ethical analysis should include four themes; 1) best standards of care and training, 2) error management and patient safety, 3) patient autonomy, and 4) social justice and resource allocation. In addition, learning from mistakes in a simulation offers opportunities to improve understanding, gain confidence, transfer knowledge and achieve appropriate performance.⁸

We report on our ongoing experience in developing, testing and evaluating the application of virtual reality simulation for medical education and training, both on-site and distributed over distance.

MATERIALS AND METHODS

These studies were developed and performed from September 2002 through May 2004. Institutional human research review boards at both University of New Mexico (UNM) and University of Hawaii (UH) approved this project for the user and student experiments. Signed informed consent was obtained from all participating subjects. The National Capital Area Simulation Center at Uniformed Services University of the Health Sciences and the Pacific Telehealth and Technology Hui at Tripler Army Medical Center and Veterans Administration, in conjunction with UH, have been participating in the ongoing planning, development and implementation of this project.

The Virtual Environment

An immersive three-dimensional (3D) environment allowed real-time exploration, examination, and manipulation of 3D objects and images.⁹ A problem-based learning (PBL)¹⁰⁻¹² case designed to demonstrate an evolving epidural hematoma in a patient (Mr. Toma) post car crash was used for these pilot studies. The interactive patient simulation allowed students to dynamically determine the outcome of the case scenario. The artificial intelligence (AI) engine was coupled to the virtual environment and represented by a virtual patient that manifested the signs and symptoms of the medical scenario. Students were fully immersed and represented within the virtual environment or observed by others from outside the virtual world. Immersed students wore a head-mounted display with trackers, allowing them a sense of presence and interaction within the virtual environment. Team members within the virtual environment were able to see each other as full human figures (avatars) and interact as if they were physically present, even when separated by significant distances. Students could examine the virtual patient, independently controlling their viewpoint and motion within the virtual world. The ratio between real time and virtual time could be varied to allow slower or faster progress of events. The immersed users worked individually or within a group to gather information and initiate interventions. When used within a PBL type tutorial, the students and tutor can discuss the case as it unfolded, pausing the scenario as appropriate, to discuss their observations, hypothesize, and generate learning issues (Figure 1). Students can also use these simulations in learning pairs or individually without a tutor.

Flatland served as the software infrastructure.¹³ It is an open source visualization/virtual reality application development environment, created at the University of New Mexico. Flatland allows software authors to construct, and users to interact with, arbitrarily complex graphical and aural representations of data and systems. It is written in C/C++



Figure 1. When used within a Problem-Based Learning (PBL) type tutorial, the students and tutor discuss the case as it unfolds, pausing the scenario as appropriate, to discuss their observations, hypothesize, and generate learning issues

and uses the standard OpenGL graphics language to produce all graphics. In addition, Flatland uses the standard libraries for window, mouse, joystick, and keyboard management. It is object oriented, multi-threaded and uses dynamically loaded libraries to build user applications in the virtual environment (VE). The end result is a virtual reality immersive environment with sight and sound, in which the operator using joy wands and virtual controls can interact with computer-generated learning scenarios that respond logically to user interaction. Virtual patients can be simulated in any of several circumstances, with any imaginable disease or injury (Figure 2).

At the core of Flatland is an open, custom, transformation graph data structure that maintains and potentially animates the geometric relationships between the objects contained in the graph. Graph objects contain all of the information necessary to draw, sound, touch, and control the entity represented by the object. By being intrinsically multi-threaded, Flatland allows the system to make use of computer systems with multiprocessors and shared memory. The main thread may spawn multiple threads to service graphics, sound, tracking and Internet-based collaboration. An application in the context of Flatland is a relatively self-contained collection of objects, functions and data that can be dynamically loaded (and unloaded) into the graph of an environment during execution. An application is responsible for creating and attaching its objects to the graph, and for supplying all object functionality. It is added to Flatland through the use of a configuration file. This structured file is read and parsed when Flatland starts, and contains the name and location of the libraries that have been created for the application, as well as a formal list of parameters and an arbitrary set of arguments for the application.

In Flatland, graphics and sound can be treated symmetrically. Sound interfaces are modeled on the OpenGL interface used for the graphics. All sound is emitted in Flatland from point sources in the 3D space. The author specifies the location of the sounds in the same model coordinate system used for the graphics.

Flatland is designed to make use of any position-tracking technology. A tracker is a multiple degree of freedom measurement device that can, in real time, monitor the position and/or orientation of multiple receiver devices in space, relative to a transmitter device. In the standard Flatland configuration, trackers are used to locate hand held wands and to track the position of the user's head. Head position and orientation are needed in cases that involve the use of head mounted displays or stereo shutter glasses (Figure 3).

User interaction is a central component of Flatland, and as such, each object is controllable in arbitrary ways defined by the designer. Currently there are four possible methods for the control of objects: 1) Pop up menus in the main viewer window, 2) the keyboard, 3) 2D control panels either in the environment or separate windows, and 4) external systems or simulations. In the future there will also be available 3D

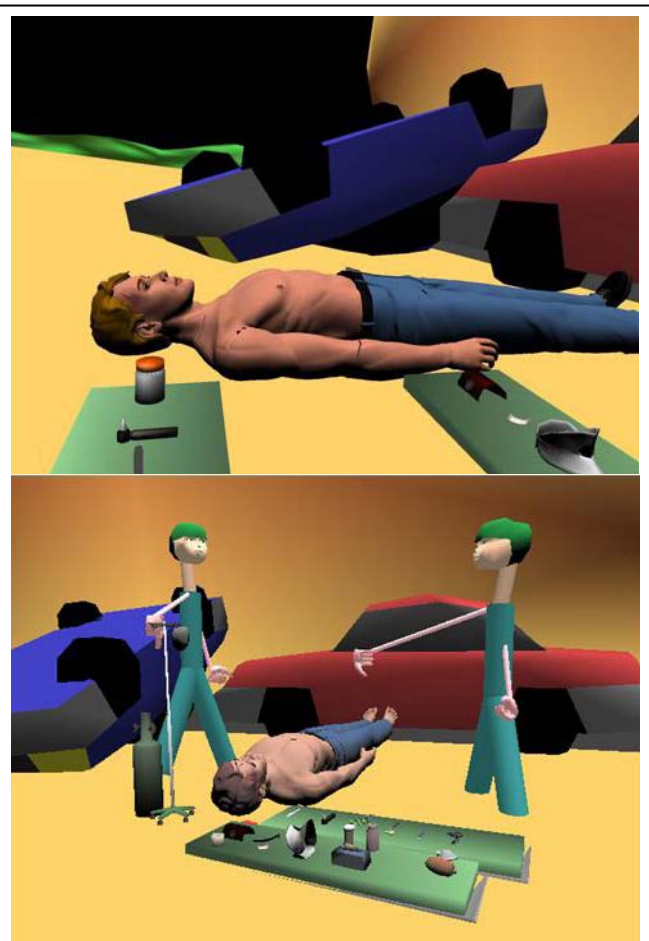


Figure 2. The simulation artificial intelligence (AI) engine dynamically governs changes in physiology, physical findings, movement and events, as well as responses to the user

menus and controls in the virtual environment and voice recognition.

The immersed user or avatar interacts with the virtual patient using a joy wand equipped with a six degree of freedom tracking system, buttons, and a trigger. The wand's representation in the environment is a virtual human hand. The user may pick up and place objects by moving the virtual hand and pulling the wand's trigger. The user avatars can also interact simultaneously as a team visually and verbally, point to point or multi-point using multi-casting over the Access Grid™ (Figure 4).^{14, 15}

Artificial Intelligence (AI)

The artificial intelligence is a forward chaining IF-THEN rule based system that specifies the behavior of objects in the VR world. The rules governing the physiology of the avatar were obtained from subject matter experts. The rules are coded in a C computer language format as logical antecedents and consequences. The AI loops over the rulebase, applying each rule's antecedents to the current state of the system, including time, and testing for logical



Figure 3. The student user virtual reality workstation

matches. Matching rules are "fired," modifying the next state of the system. Time is a special state of the system that is not directly modified by the AI, but whose rate is controlled by an adjustable clock. Since the rate of inference within the AI is controlled by this clock, the user (or student) is able to speed up, slow down, or stop the action controlled by the AI. This allows users to learn from their mistakes by repeating a scenario.

Access Grid™

The Access Grid™ (AG) is a combination of multimedia resources used to support group-to-group interactions and communication via high-speed IP/TCP networking over the Web.¹⁵ The AG allows users to share interactive experiences at multiple sites. Developed by the National Computational Science Alliance (NCSA) and led through efforts at Argonne National Laboratory, it is open source and is currently used in over 150 institutions worldwide (Figure 5).

The independent camera viewpoint application, called the Access Grid Remote Camera, captures images from the Flatland environment and transmits them over the AG for viewing at remote sites. This camera is used to capture the third-person independent view of the activities and objects within Flatland. The AG Camera can move around within



Figure 4. Students can interact and work together simultaneously in the fully immersive virtual environment independent of distance. Shown here is a student in New Mexico working in the virtual simulation with a student in Hawaii.

Flatland to any position. Multiple cameras may be launched simultaneously and separately moved for multi-view transmission into the AG. In addition, multiple participants and their avatars can be represented and tracked within the virtual environment allowing group interaction independent of distance, sharing tasks and passing off of objects. This capability permits real-time virtual team collaboration when those participants are in separate locations.¹⁴

Evaluation Methods

Evaluation consisted of several different initiatives over the past two years; 1) usability surveys by a variety of volunteers, 2) face and content validity surveys by selected subject matter experts, 3) knowledge acquisition experiments using medical students randomly selected to participate in one of four PBL-type learning formats for comparative analysis and determination of concurrent validity with a standard text-based clinical case, 4) knowledge structure relatedness non-PBL experiments using individual medical students pre and post a VR simulation experience.

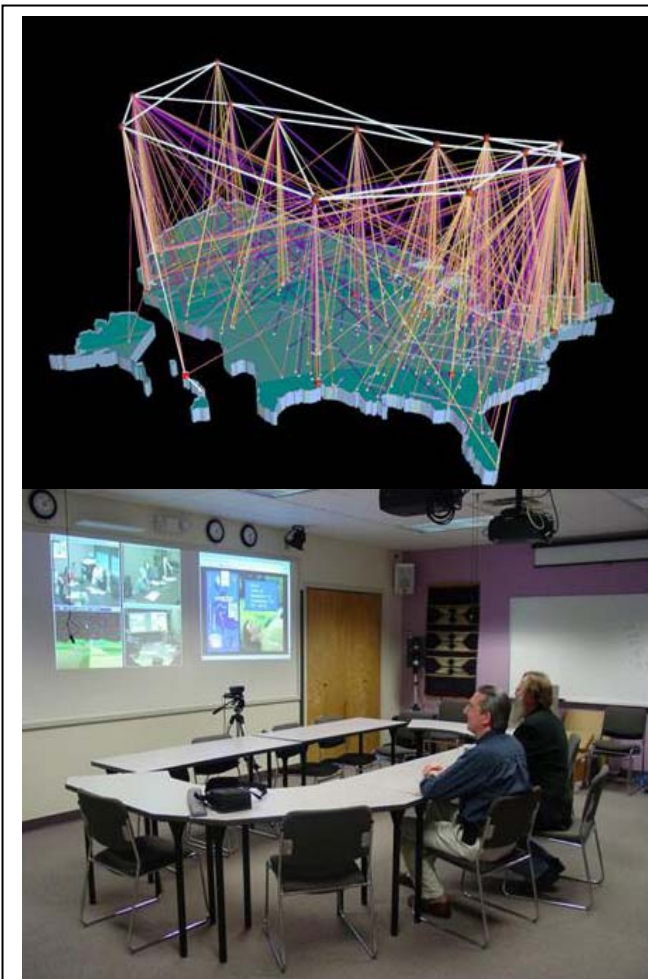


Figure 5. The Access Grid™ is an IP/TCP open source platform which supports multicasting and allows many participants to interact and multiple applications to run simultaneously

1) Usability

Usability analysis was accomplished applying a questionnaire during the virtual reality simulation experience. Demographic information sought included age, gender, health profession, area of health graduate study, experience with virtual reality, gaming, internet, and computers. There were 35 questions related to identification of objects, use of objects, comfort and VR instruction adequacy. 34 questions/statements used a one to four Lickert scale with four demonstrating highest agreement to the statement. One question asked about the number of requests for assistance.

2) Face and Content Validation

Face and content validity were determined by four subject matter experts' review of the VR simulation followed by responses to a questionnaire at the time of the review. Those subject matter experts were all medical doctors selected randomly with some experience with simulation and virtual reality, representing the fields of internal medicine, neurosurgery, neonatology/intensive care, and psychiatry.

The questionnaire was designed to obtain expert group consensus in the determination of subject groups to be used in the evaluation of learning and the educational value of the simulation. In designing the survey, consensus on the questions to present was derived from participants from previous experiments with the consideration of lessons learned from previous phases. The questionnaire itself offered an outline of the key lessons learned for the experts to consider when filling out the questions. This included issues and concerns regarding sample size, time issues and the need to obtain prior insight on the current level of competence. The questionnaire was administered via a webpage with a database backend for ease of collection and compiling of data.

3) Knowledge Acquisition and Concurrent Validation

In another set of experiments using a PBL-type format, comparative knowledge acquisition studies were done using four comparison study groups consisting of fifteen pairs of fourth year medical students and a standardized case; text-based only or VR enhanced with or without distance using the Access Grid. Knowledge acquisition by the students was determined using pre and post testing at the time of the experiments. The thirty multiple choice questions in the tests were developed by subject matter experts as they related to the concepts based on the learning goals and objectives that were integrated into the case scenario equally for all four groups. Multiple choice items were validated by an Emergency Medicine physician and a family practitioner. Items were tested for a ceiling effect using the method described by Meier (2004).¹⁶ The students were given 30 minutes to complete each test.

In order to judge short term knowledge retention, a post-post test was administered to all students approximately two weeks after their session. This test consisted of four multiple choice questions that asked the students to explain in a short written response the reason why each of the possible choices was correct or incorrect. The four questions were selected from the pre-post test. Responses classified as incorrect were assigned zero points, partially correct one point, and correct two points. All responses were graded by two faculty members blinded to the type of group in which each student participated.

4) Knowledge Structure

In a separate set of non-PBL experiments, relatedness ratings were performed using individual students. Student learning was evaluated by comparing students' knowledge structures to experts both before and after VR training. Five subject matter experts were asked to identify 25 central concepts associated with the case and the learning goals and objectives. Examples of the concepts were anisocoria, brainstem herniation and Cushing's Triad. These same experts then rated the relatedness of concept pairs and Pathfinder¹⁷ was used to derive an expert knowledge network from the averaged ratings. The 25 concepts were represented as nodes in the network and the links between nodes reflected the semantic relatedness of the concepts.

Table 1. Post Test Scores

	VR	Paper	Total
Access Grid	68.8 (5.4)	48.4 (4.4)	57.1 (4.3)
Non-distributed	57.8 (6.7)	73.4 (2.8)	65.6 (4.0)
Total	62.5 (4.5)	60.9 (4.1)	61.7 (3.0)

Scores are based on percentage correct out of 8 multiple choice questions; Mean (SE). Multiple analysis of variance with VR and Access Grid as the independent variables indicates a significant interaction between VR and Access Grid ($F = 5.22$; $df = 3, 26$; $p = 0.006$).

This expert knowledge network served as a referent against which students' knowledge networks were evaluated. Forty-eight medical students (28 males, 20 females) from the University of New Mexico and the University of Hawaii voluntarily participated in the study. They were compensated \$100 (US) for their efforts. Students ranged from their first to fourth year in medical school, with a mean of 2.96 years. Within the virtual environment, students diagnosed and treated a patient who was experiencing a hematoma as a result of a car crash. The study phase lasted approximately 1.5 hours. Both prior to and after training, participants rated the semantic relatedness of a subset of the pairs of 25 core hematoma concepts on a five-point Likert scale. Pathfinder was used to derive pre- and post-training knowledge structures for each student from the relatedness ratings. Each structure network was compared to the previously defined expert network resulting in a similarity index(s) that ranged from zero to one.

RESULTS

1) Usability

There were 21-26 respondents to each question on the usability survey. The respondent demographics consisted of 14 female and 12 male and age ranges (years); median age 40 (± 12.2 , range 19-60). Nurses comprised 11/26 (42%) of the respondents. The majority of the older age groups were in the nurse category. A declared major area of health study was stated by 13/26 respondents; Allied health (2), nursing (5), pharmacy (4), EMS (1), medicine (1). Based on a Likert scale, prior VR experience was low with 1.15/4 (range 1-2), gaming experience was slightly higher at 1.85/4 (range 1-3) on the Likert scale, with Internet experience and computer experience significantly higher at 3.54/4 (range 2-4) and 3.77/4 (range 3-4) respectively.

As part of the usability questionnaire evaluation, six pairs of questions rating identification (ID) of objects in VR vs. use of objects in VR were analyzed by applying paired T-test

using the mean score of all respondents and, if needed, the mean respondent value for any missing answers. Results of the Paired T test were $t = 4.58$, $p = .0002$, indicating the test subjects found it easier to identify objects in VR as compared to using them in VR. Chi square analysis indicates 66% of respondents found it easier to ID an object than to use the object. In addition, the average score for all identification questions (10) = 3.49/4 (range 3.00-3.77), average score of all use-related questions (18) = 2.79/4 (range 2.05-3.18). Regarding the question on number of times 23 respondents needed to ask for assistance; 0 = 11/23 (48%), 1 = 5/23 (22%), 2 = 4/23 (17%), 3 = 2/23 (9%), 4 = 1/23 (4%). 7/23 (30%) needed to ask for assistance more than once.

In the usability studies, because sample size was small, some of the variables were necessarily confounded (e.g. nurse and age groupings). Thus, the interpretation of correlations on these variables must be qualified.

Comfort of the head mounted display was rated by 26 respondents with an average score of 2.54/4, with the following comments from individual respondents; discomfort (4), hard to use with glasses (3), loose fit (3), heavy (1). Lack of ill effects of immersion was scored 2.88/4 with the following comments from individual respondents; nausea (3), mild dizziness (2), motion sickness (1), eye strain (1), disorientation and off-balance (1). Usability studies allowed improvements in user interface, locomotion and navigation.

2) Face and Content Validation

Upon review of the case-based simulation, the four subject matter experts found face and content validity in the content and representations in the simulation based on the learning goals and objectives of the case. Subsequently, after all the experts completed the questionnaire, the data was gathered, organized and posted on a webpage for experts to discuss and review. The results were available for review previous to meeting so that the experts had a chance to review their own answers, as well as the other participants. A general meeting was held to review all the results and obtain a consensus on how to proceed. The questionnaire enabled the experts to individually assemble and record their opinions, ideas, issues and concerns and later view the total results for discussion and conclusions.

3) Knowledge Acquisition

In these experiments, fifteen pairs of medical students in Hawaii and New Mexico participated collaboratively in problem solving and management using a text-based case or a simulated patient in VR on site or independent of distance over the Access Grid. Some students indicated that the "ability to interact with a colleague from a distance is helpful" and that better understanding of the concepts in the case came from "being able to communicate with a colleague". Student reaction provided confirming evidence that distributed learning enables interchange between geographically distant students and allows students from different institutions to interact. VR did create higher

Table 2. Gain in Scores from Pre-Test to Post Test

	VR	Pape	Total
Access Grid	29.2 (6.2)	20.3 (5.2)	24.1 (4.0)
Non-distributed	20.3 (4.7)	40.6 (3.1)	30.5 (3.8)
Total	24.1 (3.8)	30.5 (3.9)	27.5 (2.8)

Scores are based on percentage correct out of 8 multiple choice questions; Mean (SE). Multiple analysis of variance with VR and Access Grid as the independent variables indicates a significant interaction between VR and Access Grid ($F = 4.31$; $df = 3, 26$; $p = 0.01$).

performance expectations and some anxiety among VR users. VR orientation was reported as adequate but students needed time to adapt and practice in order to improve efficiency. Students who used VR stated they felt more engaged than when using text-based cases. Students stated that opportunities to make mistakes and repeat actions in the VR were extremely helpful in learning specific principles. Post-testing performance was similar between the VR and non-VR groups (see Table 1), indicating VR was not a detractor to the learning experience. A 31% average knowledge gain of students participating in non-distributed sessions compared with a 24% average knowledge gain of students in distributed sessions indicated distance was not a barrier to the learning experience (see Table 2). The non-VR non-distributed students had the highest knowledge gain (41%). However, the VR non-distributed and non-VR distributed showed the lowest knowledge gain (20%). Multiple analysis of variance of the gain in scores on the post test using VR and AG as independent variables indicated a significant interaction between those two variables ($p = 0.01$). Multiple analysis of variance of post-post test scores revealed no significant differences among the 4 groups using VR and AG as independent variables (see Table 3). None of the experimental conditions led to any knowledge loss or misunderstanding of key concepts. The similarity in learning acquisition with or without distance and with or without VR demonstrates concurrent validity with the current PBL case text-based approach.

4) Knowledge structure

In this set of separate experiments there were 48 students who completed the training including both sets of relatedness ratings. There were 28 males and 20 females. Students ranged from first year to fourth year in their programs with the mean number of years equaling 2.96.

Students rated the relatedness of 72 pairs of concepts critical to the case, 36 of which were related as defined by an expert knowledge network, and 36 unrelated. The 36 unrelated

Table 3. Post Post Test Scores

	VR	Paper	Total
Access Grid	19.8 (9.3) range = 9 - 32	22.4 (5.6) range = 16 - 32	21.2 (7.3)
Non-distributed	24.3 (7.8) range = 14 - 32	21.9 (3.5) range = 17 - 26	23.1 (6.0)
Total	22.4 (8.5)	22.1 (4.4)	22.2 (6.6)

Scores are based on one point for each partially correct explanation and two points for each correct explanation on the 19 essay questions. Average scores for each group and standard deviations (in parentheses) are reported. Multiple analysis of variance with VR and Access Grid as the independent variables: $F = 0.50$; $df = 3, 25$; $p = 0.68$.

pairs were used primarily as foils to balance the related pairs. First each student's raw ratings were correlated with the expert ratings. The average correlation was $r = 0.77$ for pre training and $r = 0.76$ for post training, both highly statistically significant ($p < 0.001$). The high correlation at pre training indicates that the students were knowledgeable of the terms and their relationships even before training. This conclusion is also corroborated by comparing the mean rating of the related pairs (4.25) with the mean rating of the unrelated pairs (1.96) on the set of relatedness ratings before training.

Pathfinder was then used on each of the students' raw ratings to derive a knowledge network. Each student's knowledge network was compared to the expert knowledge network using a method that produces a similarity index (s) that varies from zero to one.

The mean similarity scores of all student subjects ($N = 48$) were $s = 0.70$ before training and $s = 0.72$ after training. A matched t-test on these differences resulted in $t = 1.172$, $p = 0.247$. Although the difference did not reach statistical significance, it was trending toward statistical significance (see Table 4).

The pre/post difference may have failed to reach significance because of a ceiling effect; students were already performing at a high level before training. To assess this, a cutoff value was used $s = .80$ to select the subset of students who had similarity to expert scores below this level on the pretest knowledge structures. This resulted in selecting a subset of 36 out of the 48 students. A matched pairs t-test for these 36 students was performed on the difference between the similarity to expert scores before and after training and found $t = 2.577$, $p = .014$; a highly statistically significant difference (see Table 4).

Table 4. Pre and Post Knowledge Structure Similarity Scores

	Pre Training	Post Training	paired t-test
N ¹ =48	0.70 (0.097)	0.72 (0.098)	t = 1.172, p = 0.247
N ² =36	0.66 (0.077)	0.70 (0.096)	t = 2.577, p = 0.014

The mean similarity scores comparing students' to experts' knowledge structure relatedness pre and post virtual reality training. N¹ are the similarity scores of all student participants. N² are the sub-set of students with a similarity score of < 0.80 pre-training. Mean (SE)

DISCUSSION

Although the smaller sample size and resulting confounding variables in the usability analysis prevented sub-sample analysis, the group as a whole rated identification of objects in VR significantly higher than use of those objects in VR (paired T test). This may indicate the graphics of VR objects are satisfactory but manipulation of those objects is more difficult and deserves review and improvement or the need for better training to competence in using those tools. In the future, this usability component of our studies would be better designed by increasing the sample size of participants to reflect the potential heterogeneity of the user groups or applying prospective controlled selection criteria to insure a desired balanced mix of potential users.

Although there were no significant overall differences in knowledge acquisition among the different student groups, the quality of the VR or distance learning experience equaled that of the traditional PBL case and there was no knowledge loss. These findings indicate VR or distance did no harm in the learning experiences related to this case, demonstrates concurrent validity with the traditional PBL format, and offers evidence that these methods can be used on-site and in a distributed manner. The participant's ability to act in an environment experimentally, and not just observe it, is a critical feature that simulation provides. According to Winn,¹⁸ the theoretical assumption of learning from simulation is that students can construct understanding for themselves by interacting with information and materials—an orientation to learning that has acquired the name “constructivism.”¹⁹ Winn makes a distinction between simulation and reification. The purpose of simulation is to represent real-world objects in as accurate a way as possible. Reification “is the process whereby phenomena that cannot be directly perceived and experienced in the real world are given qualities of concrete objects that can be perceived and interacted within a virtual learning environment.”²⁰ Currently we are applying the reification concept to development of a renal and nephron model with which the learner can interact in VR in order to enhance understanding of a variety of renal physiologic concepts, such as the counter-current concentration mechanism, as well as pharmacologic or pathophysiologic effects on the reified model.

Of central importance to our proposed work is the ability to assess how well someone has learned a complex, conceptually demanding area. Our hypothesis that simulation is better than conventional learning environments needs to be evaluated with psychometrically sound assessments. One accepted method is the use of knowledge structure. Over the past decades, an impressive literature has accumulated showing that the structural properties of domain knowledge are closely related to domain competence.²¹ How someone has the central concepts of an area organized in memory relates to his or her level of knowledge. Experts share a particular structural organization of concepts, and as a consequence, are more likely to see certain relevant abstract relationships and connections.²² Studies have shown that experts in various domains, such as physics and computer programming, organize the central concepts along semantic dimensions, whereas novices focus on surface level characteristics.²³ It is this organization of domain knowledge, we believe, that reflects an individual's degree of conceptual understanding in a domain. Structural approaches to assessing domain knowledge began to appear in the late 1960s and early 1970s.^{24, 25} Several investigators reported finding that classroom performance was related to students' structural organization of the central concepts in a course. In our own work, we focused on developing a systematic three-phase methodology for implementing structural assessment (SA): (a) Elicitation: evoking some behavioral index of an individual's organization of domain concepts, (b) Representation: applying statistical techniques to transform the elicited data into a formal representation (e.g., network) that captures the important structural properties of the knowledge, (c) Evaluation: quantifying the level of expertise that is reflected in a derived representation.

The elicitation phase comprises three steps: (1) defining core concepts in the domain by rank-ordering experts' ratings of a list of domain concepts drawn from textbooks and expert input; (2) defining the expert structure based on experts' ratings of all pair wise combinations of concepts; and (3) eliciting proximity data by presenting participants with a target concept from the set of most related pairs in the expert structure accompanied by three choices (one of which is the most related pair item). And finally, having them choose the concept they believe is most related. The viability of this elicitation procedure rests on the assumption that relatedness ratings on pairs of concepts are a valid and reliable measure

of individuals' semantic structure. There exists a long history of research and theory going back to James²⁶ and continuing with the work of Shepard²⁷ and Tversky²⁸ that supports this approach.

Representation entails finding a method that converts a proximity matrix of raw relatedness ratings into a form that best elucidates the underlying structure of the relations. The resulting representation should: (a) capture the structural relations among concepts; (b) be easy to comprehend; (c) capture all relevant (e.g., predictive) latent structure; and (d) be data-driven. Several different scaling algorithms have been evaluated—including multidimensional scaling,²⁹ hierarchical clustering,³⁰ and untransformed proximity data—and it was found that the Pathfinder¹⁷ algorithm best met the above criteria.^{31, 32} Pathfinder generates a connected graph that depicts local concept relationships, which according to Latour,³³ provide the most compact and powerful way of representing data. Most importantly, Pathfinder was the most predictive representation of classroom performance.

Finally, a participant's knowledge structure must be evaluated in terms of the level of competence or sophistication it represents. Two methods were used for accomplishing this. A referent-based evaluation compares a student's Pathfinder network to an expert Pathfinder network, resulting in an index of similarity between 0 and 1.³⁴ Evaluating a family of similarity indices, indicated that similarity based on the commonality of directly linked concepts provided the best predictor of student performance.³¹ This measure of structural similarity is referred to as SIM. A referent-free method of evaluating an individual's knowledge structure was also developed. This method is called "coherence." Coherence measures (0-1 scale) the internal consistency of a set of ratings by examining the extent to which sets of ratings satisfy a generalized triangle inequality law. It has been found that coherence increases with levels of domain expertise.³⁵

Using these approaches in the knowledge structure experiments of our studies, students knew the central concepts and their relations for the hematoma case fairly well before training. The data are fairly convincing in support that a ceiling effect was operating in the knowledge structure experiments. We report that the correlation between students' and expert ratings were $r=.77$ (highly significant) even before any training and that there was a large difference in the mean ratings of the related concept pairs (4.25) and unrelated concept pairs (1.96) again before any training took place. These two findings strongly suggest that students were highly knowledgeable of the concepts and their relations to one another before training. By selecting a subset of students who scored below $s=.80$ on the pretest and then finding that with these students selected only on the basis of how much knowledge they had before training, a significant increase in similarity to expert knowledge structure after training argues for a ceiling effect in the larger group of students. By examining those students (36/48) who were lower in their knowledge of the concepts

before training, we find that their knowledge structures after training are in fact closer to the expert knowledge network than before training, a statistically significant change in that sub-set. There is a modest but statistically significant improvement in students' understanding as a function of the training experience. Two possibilities include the VR training itself and the ancillary material provided. Designing experiments by selecting cases with or without ancillary materials and students with lower levels of correlation with expert knowledge structure may assist in evaluating the impact of the simulation experience on learning. There remains a need for more validation, training to competence in using the VR tools, evaluation of learning impact, knowledge transfer and effects on performance.

Currently, enhancements of simulations with integration of sound and haptics into the VR environment are being explored. The haptics function within an object contains all of the calls or code to produce force or tactile sensations. Without haptics, a user who reaches out to touch a virtual reality object discovers his hand will move through the object, which can be disconcerting. With haptics, the user experiences force feedback. The result is the sensation of actual touch. Currently, only one haptic device is supported in Flatland, and that is called the PhantomTM. We are also beginning to integrate sound into the virtual environment to enhance the sense of presence and reality in the virtual experience. Examples of sound integration include the ability for auscultation with the virtual stethoscope to hear breath and bowel sounds, audible responses of the virtual patient to stimulus or pain, sound elicited by the use of certain virtual tools, and ambient noises. Based on our usability studies, we will be remodeling some of the interactive, locomotion and navigation metaphors and tools used within the Flatland virtual environment in order to further improve the user interface, as well as optimize the actual learning or training experience. In conjunction with the Uniformed Services University, we are also porting Flatland to a Windows environment in order to increase the potential user base and compatibility with other systems.

CONCLUSIONS

This research cuts across the integration of computing, networking, human-computer interfaces, learning, and knowledge acquisition. VR creates a safe environment to make mistakes and could allow rapid deployment for just-in-time training or performance assessment.³⁶⁻³⁹ These experiments have demonstrated virtual collaboration within VR is possible with multiple participants independent of distance. Students accept use of VR for education and training. Participants stated they felt more engaged in VR. Students also felt they learned best from their mistakes in VR. In comparative experiments, post-testing performance was similar between VR and non-VR Groups, as well as distributed and non-distributed groups, indicating VR or distance distribution "does no harm" and demonstrating concurrent validity with standard text-based problem-based learning (PBL) case methods.

Perhaps most significant was the evidence of improvement in knowledge structure after the virtual reality simulation experience in a select group of learners with initial lower levels of knowledge structure correlation with that of experts. Knowledge structure relatedness ratings were significantly improved in those students with lower pre-VR relatedness ratings which indicate the potential value of simulation in learning, particularly in those students with lower levels of knowledge structure as related to the concepts being learned. In general, the results suggest that the method used for eliciting, representing and evaluating knowledge structure offers a sensitive, objective and valid means for determining learning in virtual environments. More research is indicated using controlled studies to determine more specifically what aspects of the learning event produced the changes in knowledge.

The initial research and development of our virtual reality simulation required significant time and resources through an extensive iterative process requiring ongoing evaluation and improvements. Although our studies were not designed to evaluate cost-benefit, we anticipate that with continued experience, production time and cost should diminish. In addition, we speculate that by using these simulations for distributed team training travel costs and time would be avoided and consistency in the training scenarios could be better achieved. We now plan to develop more simulations using a knowledge-based design approach that would be validated for learning and training, as well as evaluated for impact on learning and performance. We envision creation of a library of simulations that can be integrated into curricula and training programs, modified to meet specific learning goals and objectives of a variety of learners or trainees. Further, a production algorithm is being developed to allow more rapid creation of simulations on demand. A protocol has been developed for communication between subject matter experts and programmers that allows incorporation of medical knowledge into the VR system rapidly. That includes behavioral, physiological, medical and other content-related features of the simulation. We have developed an effective technical pipeline from 3D authoring software to the VR system. We use Maya™ by Alias for creating shapes, textures and animations. To incorporate 3D art created in Maya into Flatland takes a few minutes and the production cycle iterates quickly, allowing incremental improvements of the look-and-feel of the 3D content, until finally approved by subject matter experts. Similar "Maya-to-game" data transition is being used by the leading computer gaming companies in a time efficient manner.

We anticipate development of more collaborative team learning and training initiatives, independent of distance, allowing participants to be physically separated but virtually together. These simulations also provide opportunities for multidisciplinary education. And, in addition to education, they may be useful for screening or testing applications. These methods can create "Just-in-Time" training potential, performance assessment platforms, and more national and international collaborative opportunities.

REFERENCES

1. Dawson-Saunders, B., Feltovich, P.J., Coulson, R.L. and Steward, D.E. A survey of medical school teachers to identify basic biomedical concepts medical students should understand. *Academic Medicine*. 1990; (65): 7: 448-454.
2. Windish, D.M., Paulman, P.M., Goroll, A.H. and Bass, E.B. Do clerkship directors think medical students are prepared for the clerkship years? *Academic Medicine*. 2004; 79: 56-61.
3. Kohn, L., Corrigan, J.M. and Donaldson, M. To err is human: Building a safer health system. Committee on Quality of Health Care in America. Institute of Medicine. *Washington, D.C. National Academy Press*, 1999.
4. Committee on Quality of Health care in America. Crossing the quality chasm: A new health system for the 21st century. Institute of Medicine, *Washington, D.C. National Academy Press*, 2001.
5. Champion, H.R. and Higgins, G.A. Meta-Analysis and Planning of SIMTRAUMA: Medical Simulation for Combat Trauma Training. *USAMRMC TATRC Report*. 2000; No.00-03.
6. Satava, R.M. and Jones, S.B. The Future is Now: Virtual Reality Technologies; Innovative Simulations for Assessing Professional Competence: from Paper-and-Pencil to Virtual Reality. Tekian A, McGuire CH, McGaghie WC and Associates. *University of Illinois at Chicago, Department of Medical Education*. 1999; (12): 179-193 .
7. Ziv, A., Wolpe, P.R., Small, S.D. and Glick S. Simulation-Based Medical Education: An Ethical Imperative. *Academic Medicine*. 2003; 78(8): 783-788.
8. Alverson D., Saiki, S.M., Jacobs, J., Saland, L., Keep, M.F., Norenberg, J., Baker, R., Nakatsu, C., Kalishman, S., Lindberg, M., Wax, D., Mowafi, M., Summers, K.L., Holten, J.R., Greenfield, J.A., Aalseth, E., Nickles, D., Sherstyuk, A., Haines, K. and Caudell, T.P. Distributed interactive virtual environments for collaborative experiential learning and training independent of distance over Internet 2; Medicine Meets Virtual Reality 12; Building a Better You: The Next Tools for Medical Education, Diagnosis, and Care, Volume 98. *Studies in Health Technology and Informatics*. Westwood, J.D., Haluck, R.S. and Amsterdam, H. *Amsterdam, the Netherlands: IOS Press*; 2004; 98: 7-12.
9. Jacobs, J., Caudell, T., Wilks, D., Keep, M.F., Mitchell, S., Buchanan, H., Saland, L., Rosenheimer, J., Lozanoff, B.K., Lozanoff, S., Saiki, S. and Alverson, D. Integration of Advanced Technologies to Enhance Problem-Based Learning over Distance: Project TOUCH. *Anatomical Record*. 2003; 270B: 16-22 .
10. Kaufman, A., Mennin, S., Waterman, R., Duban, S., Hansbarger, C., Silverblatt, H., Obenshain, S., Kantrowitz, M., Becker, T., Samet, J. and Wiese, W. The New Mexico experiment: educational innovation and institutional change. *Academic Medicine*. 1989; 64: 285-294.

11. Anderson, A. Conversion to problem-based learning in 15 months; The Challenge of Problem Based Learning. D. Boud and G. Feletti. *New York, NY. St. Martin's Press*, 1991; 72-79.
12. Bereiter, C. and Scardamalia, M. Commentary on Part I: Process and product in problem-based learning (PBL) research; Problem-based learning: A research perspective on learning interactions. D. H. Evensen and C. E. Hmelo. *New Jersey. Lawrence Erlbaum Assoc. Publishers*, 2000; 185-195.
13. Caudell, T.P., Summers, K.L., Holten, J., Takeshi, H., Mowafi, M., Jacobs, J., Lozanoff, B.K., Lozanoff, S., Wilks, D., Keep, M.F., Saiki, S. and Alverson, D. A Virtual Patient Simulator for Distributed Collaborative Medical Education. *Anatomical Record*. 2003; 270B: 16-22.
14. Mowafi, M., Summers, K.L., Holten, J., Greenfield, J.A., Sherstyuk, A., Nickles, D., Aalseth, E., Takamiya, W., Saiki, S., Alverson, D. and Caudell, T.P. Distributed interactive virtual environments for collaborative medical education and training: Design and characterization; J.D. Westwood, R.S. Haluck and H. Amsterdam. *Medicine Meets Virtual Reality 12; Building a Better You: The Next Tools for Medical Education, Diagnosis, and Care*, Volume 98. *Studies in Health Technology and Informatics. Amsterdam, the Netherlands: IOS Press*, 2004; 98: 259-261.
15. Childers, L., Disz, T.L., Hereld, M., Hudson R., Judson, I., Olson, R., Papka, M.E., Paris, J., and Stevens, R. Active Spaces on the Grid: The Construction of Advanced Visualization and Interaction Environments; B. Engquist, L. Johnsson, M. Hammill, and F. Short. *Paralleldatorcentrum Kungl Tekniska Högskolan Seventh Annual Conference (Simulation and Visualization on the Grid)*, vol. 13, *Lecture Notes in Computational Science and Engineering. Stockholm, Sweden: Springer-Verlag*, 1999; 64-80.
16. Meier, S. Improving Design Sensitivity through Intervention-Sensitive Measures. *American Journal of Evaluation*. 25(3): 321-334.
17. Schvaneveldt, R.W. Pathfinder Associative Networks: Studies in Knowledge Organization. *Norwood, NJ. Ablex*, 1990.
18. Winn, W.D. Current trends in educational technology research: The study of learning environments. *Educational Psychology Review*. 2002; 14(3): 331-351.
19. Duffy, T. and Jonassen, D. Constructivism and the technology of instruction: A conversation. *Hillsdale, NJ: Lawrence Erlbaum Association*, 1992.
20. Winn, W.D. A conceptual basis for educational applications of virtual reality. *Human Interface Technology Laboratory Technical Report TR-93-9. Seattle, WA: Human Interface Technology Laboratory, University of Washington*, 1993.
21. Glasser, R. On the nature of expertise; F. Klix and H. Hagendorf. *Human Memory and Cognitive Capabilities: Mechanism and Performances. North Holland: Elsevier Science*, 1986.
22. Bransford, J., Brown, A.L. and Cocking, R.R. *How People Learn: Brain, Mind, Experience, and School. Washington D.C. National Academy Press*, 2000.
23. Chi, M.T.H., Glaser R. and Rees, E. Expertise in Problem Solving; R.J. Sternberg. *Advances in Development of Human Intelligence*, Vol. 1. *Hillsdale, NJ: Lawrence Erlbaum Associates*, 1982.
24. Geeslin, W.E. and Shavelson, R.J. Comparison of content structure and cognitive structure in high school students learning of probability. *Journal of Research in Mathematics Education*. 1975; 6: 109-120.
25. Shavelson, R.J. and Staton, G.C. Construct validation: Methodology and application to three measures of cognitive structure. *Journal of Educational Measurement*. 1975; 12: 67-85.
26. James, W. *The Principles of Psychology. Cambridge, MA: Harvard University Press*, 1890/1983: 434.
27. Shepard, R.N. Toward a universal law of generalization for psychological science. *Science*. 1987; 237: 1317-1323.
28. Tversky, A. Features of similarity. *Psychological Review*. 1977; 84: 327-352.
29. Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*. 1964; 29: 1-27.
30. Johnson, S.C. Hierarchical clustering schemes. *Psychometrika*. 1967; 32: 241-254.
31. Johnson, P.J., Goldsmith, T.E. and Teague, K.W. Structural knowledge assessment: Locus of the predictive advantage in Pathfinder-based structures. *Journal of Educational Psychology*. 1994; 86: 617-626.
32. Johnson, P.J., Goldsmith, T.E. and Teague, K.W. Similarity, structure, and knowledge: A representational approach to assessment; Nichols, Chipman and Brennan. *Cognitively Diagnostic Assessment. Hillsdale, NJ: Lawrence Erlbaum Associates*, 1995: 221-249.
33. Latour, B. Drawing things together; M. Lynch and S. Woolgar. *Representation in Scientific Practice. Cambridge, MA: MIT Press*, 1990: 19-68.
34. Goldsmith, T. and Davenport, D. Assessing structural similarity of graphs; R. Schvaneveldt. *Pathfinder Associative Networks: Studies in Knowledge Organization. Norwood NJ: Ablex Publishing Corporation*, 1990.
35. Acton, W.H., Johnson, P.J. and Goldsmith, T.E. Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology*. 1994; 85: 88-96.
36. Kolb, D.A. *Experiential Learning: Experience as the Source of Learning and Development. Upper Saddle River, NJ: Prentice Hall*, 1983.
37. Issenberg, S.B. and McGaghie, W.C. Assessing Knowledge and Skills in the Health Profession: A Continuum of Simulation Fidelity; A. Tekian, C.H. McGuire, W.C. McGaghie and Associates. *Innovative Simulations for Assessing Professional Competence: from Paper-and-Pencil to Virtual Reality. University of Illinois at Chicago, Department of Medical Education*, 1999: 125-146.
38. Issenberg, S.B., McGaghie, W.C., Hart, I.R., Mayer, J.W., Felner, J.M., Petrusa, E.R., Waugh, R.A., Brown, D.D., Safford, R.R., Gessner, I.H., Gordon, D.L. and

Ewy, G.A. Simulation Technology for Health Care Professional Skills Training and Assessment. *JAMA*. 1999; 282: 861-866.

39. Issenberg, S.B., McGaghie, W.C., Gordon, D., Symes, S., Petrusa, E.R., Hart, I.R. and Harden, R.M. Effectiveness of a Cardiology Review Course for Internal Medicine Residents Using Simulation Technology and Deliberate Practice. *Teaching and Learning in Medicine*. 2002; 14(4): 223-228.

ACKNOWLEDGEMENTS

The project described was supported partially by grant 2 D1B TM 00003-02 from the Office for the Advancement of Telehealth, Health Resources and Services Administration, Department of Health and Human Services. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Health Resources and Services Administration.