

## **Applying Virtual Reality in Medical Communication Education: Current Findings and Potential Teaching and Learning Benefits of Immersive Virtual Patients**

Benjamin Lok<sup>1</sup>, Richard E. Ferdig<sup>2</sup>, Andrew Raij<sup>1</sup>, Kyle Johnsen<sup>1</sup>, Robert Dickerson<sup>1</sup>,  
Jade Coutts<sup>2</sup>, Amy Stevens<sup>3</sup>, D. Scott Lind<sup>4</sup>

University of Florida, College of Engineering<sup>1</sup>  
University of Florida, College of Education<sup>2</sup>  
University of Florida, College of Medicine<sup>3</sup>  
Medical College of Georgia<sup>4</sup>

The purpose of this paper is to summarize a theoretical perspective toward the development of a virtual reality innovation in education. Next, we will describe a virtual character project that is impacting the training of medical students in two institutions. In doing so, we will present a summary of three studies completed over the last two years, highlighting specific results from the research. Finally, we conclude with a discussion of the impact of these findings on the development and implementation of virtual reality systems for teaching and learning.

It is our hope that through providing this study of a virtual reality scenario - experienced by over one hundred end-users - from an educational concepts perspective will help others aiming to apply virtual reality to education.

### **A theoretical perspective on VR and Virtual Character development**

Although research suggests various, promising findings for technology innovations like Virtual Reality for teaching and learning, all good innovations must start with good pedagogy (Ferdig, 2005). From a social constructivist perspective, this means:

- Virtual Reality Innovations must be imbued with authentic, interesting, and challenging academic content (at the high end of the students' Zone of Proximal Development).
- Participants must have a sense of ownership.
- There must be opportunities for active participation and social interaction.
- VR must provide chances for the creation of artifacts in a variety of ways.
- Publication, reflection, and feedback play a key role throughout the Virtual Reality Tool.

*Authentic, interesting, and challenging content.* Authentic content refers to content that is meaningful and anchored in a real-world problem (Newman, Secada, & Wehlage, 1995). Albanese states that this type of learning is an instructional methodology characterized by the use of problems as a context for students to learn problem-solving skills and acquire knowledge about the topic they are studying (Albanese & Mitchell, 1993). It is important to have authentic, real-world problems because they are interesting and meaningful to the students and thus engaging. Interesting problems, in turn, create significant missions for the students to fulfill; learning occurs in the context of carrying out that mission (Kolodner, 1997).

Along with being authentic and interesting, content that is supported by technology must be challenging to the students. A main tenet of Vygotsky's theory is the

importance of aiming instruction at the upper boundaries of a student's 'Zone of Proximal Development' or 'ZPD' (Brown & Ferrara, 1985). The ZPD is defined as: the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers. (Vygotsky, 1978, p.86)

In other words, if instruction is too easy for the user, they will lose interest; if it is too hard, they will become frustrated. The goal is to use content that is at the high end of their ZPD, where learning takes place with adult guidance or collaboration with more knowledgeable or more capable others. The student still acts as the agent in the learning activities, but knowledge emerges from the social interactions between the student and the more knowledgeable other (Scardamalia & Bereiter, 1991). These other participants scaffold the learning such that the individual constructs knowledge at a level unreachable by him or herself alone.

*A Sense of Ownership.* The active construction of knowledge means that the student learns to take on a self-regulating role in the learning process. This active construction has become the forefront of many education mission statements, specifically stating: "the self-regulated learner must have a healthy self-concept with a strong understanding that they, alone, are in control of their learning, mastery of tasks, and attainment of goals" (Sandford & Richardson, 1997). The emphasis is on student control of their learning, where opportunities for that ownership are available in the design as well as the solution of the project or problem. Technologies like virtual reality can offer ways for students to establish that personal intellectual ownership of new concepts while they visualize and interact with abstract ideas (O'Shea, 1999).

*Active Participation and Social Interaction.* Closely tied to the idea of the Zone of Proximal Development is the notion that VR must provide opportunities for active participation, collaboration and social interaction. Active participation has seemingly become a catch phrase in any learning theory that opposes itself to "traditional didactic approaches to education, which seem to be based on an assumption of direct transfer of knowledge from teacher to student, without an intervening constructive process" (Scardamalia & Bereiter, 1991, p. 38). In other words, knowledge is not transmitted from the expert to a passive learner; rather, learning is an enculturation process where knowledge is actively constructed within the student's ZPD with the help of more capable others (Brown, Collins, & Duguid, 1989; Rogoff, 1994).

Regardless of who the more capable other is, technology can support the active construction of knowledge and eventually the taking over of the self-regulating role in the social learning relationship. Innovations that espouse active learning, collaboration, and social interaction also offer opportunities for new types of relationships between teachers and students—least of which is the proverbial move from 'sage on the stage' to the 'guide on the side' (Batson, 1993). Finally, innovations become promising tools inasmuch as they provide space for the creation of learning communities (Lave & Wenger, 1991). Those communities, places where students can try out ideas and challenge the ideas of others, are both supported through and emergent from interactions with technology such as computers (Krajcik et al., 1994).

*The Creation of Artifacts.* Michael Cole (Cole, 1996) states: "an artifact is an aspect of the material world that has been modified over the history of its incorporation into

goal-directed human action" (p. 117). In social constructivist thought, these artifacts are integral and inseparable components of human functioning (Engestrom, 1991). The creation of those artifacts allows students to learn concepts, apply information, and represent knowledge in a variety of ways (Blumenfeld et al., 1994). Those artifacts, in turn, represent students' understanding of the problem, resulting solutions, and emergent states of knowledge (Krajcik et al., 1994). Virtual Reality environments must provide opportunities for students not just to passively experience, but also to create artifacts of that experience in the process of learning.

*Publication, Reflection, and Feedback.* A final critical component is the opportunity for users of VR innovations to publish, reflect, and receive feedback on their efforts. This is essential to a social constructivist model of learning because of what Rom Harré (Harré, 1984; Harré, Clarke, & DeCarlo, 1985) has called the 'Vygotsky Space.' His representation helps clarify how learners "move from using new meanings or strategies publicly and in interaction with others to individually appropriating and transforming these concepts and strategies into newly invented ways of thinking" (Gavelek & Raphael, 1996). The Vygotsky Space defines and describes four recursive processes within the individual-social and public-private dimensions: appropriation, transformation, publication, and conventionalization.

Publication is the process in which student knowledge, understandings and strategies are made public so that others can respond. Artifact creation and the opportunity for publication are important ingredients in good innovations for three reasons. First, through publications, teachers and researchers "can infer the process by which students transform meanings and strategies appropriated within the social domain, making those strategies their own" (Gavelek & Raphael, 1996) p. 188). Second, publishing makes material accessible to subsequent reflection and analysis, allowing students to revisit and revise their artifacts, thus enriching the learning experience (Krajcik et al., 1994).

A third reason publication is important refers back to the need for a good innovation to consist of challenging, academic content at the high end of the Zone of Proximal Development. Assistance from a more capable or more knowledgeable other in the ZPD is referred to as scaffolding (Wood, Bruner, & Ross, 1976). "Scaffolding characterizes the social interaction that occurs among students and teachers that precedes internalization of the knowledge, skills, and dispositions useful for all learners" (Roehler & Cantlon, 1996). Publication offers the opportunity for feedback; feedback, in turn, scaffolds a learner in their quests for knowledge construction, knowledge integration (Linn, 1991), higher-order thinking, and self-regulatory behavior.

### **Current efforts in Virtual Reality and Virtual Characters**

Good pedagogy guides good virtual reality development. There are several examples where researchers have used pedagogical principals and developed training, teaching, and learning environments. Thórisson (1997) presented an interactive guide named Gandalf that takes users on tours of the solar system. USC's Institute for Creative Technologies has created virtual experiences to train military personnel in interpersonal leadership (Hill et al., 2003). The Just VR system (Manganas et al., 2004) allows a medical trainee to interact with a virtual assistant to assess and treat a virtual victim. The Human Modeling and Simulation Group at the University of Pennsylvania uses virtual humans for task analysis and assembly validation (Badler, Erignac, & Liu, 2002). Pertaub, Slater, & Barker (2001) observed participants with a

fear of public speaking, speaking to an audience of virtual characters. They responded similarly to when they spoke to an audience of real people; further, they found that experiencing a virtual social situation may reduce anxiety in reality. Garau et al. (2001) showed that realistic, task-appropriate avatar eye-gaze behavior led to improved communication between the people represented by the avatars. Bailenson et al (2001; 2004) have shown that people manage their personal space when interacting with virtual humans similarly to when they interact with real humans. They found that people displayed a tendency to put more space between them and an embodied tutor than they did with strangers (2004) and participants maintained more distance from embodied agents than inanimate virtual objects (2001). Female participants maintained more distance from embodied agents that maintained eye contact than with agents that did not.

Advances in rendering, audio, and animation allow virtual humans to be presented with increasing levels of fidelity. Improvements in tracking, gesture recognition, and voice recognition also enable natural means of interaction. This combination of high-fidelity output and natural input has led to research into the use of virtual humans as partners in interpersonal scenarios. The concept of interpersonal, virtual humans raises an important question: How is experiencing an interpersonal scenario with a virtual character similar to – and different from – experiencing one with a real person? Clearly there must be differences, as no one would be “fooled” by a virtual character into thinking they were interacting with a real person. But, in which ways can they be similar? What are the key differences?

We have found little work that directly compares real and simulated interpersonal scenarios. However, researchers have compared other virtual environments to their real counterparts. In the psychology domain, Emmelkamp et al. (2002) compared the reactions of acrophobes in similar virtual and real environments. Using standardized measures of acrophobia, the authors found that exposure therapy in the virtual environment was as effective as therapy in the real environment. Rothbaum et al. (2000) compared virtual and real exposure therapy for those with fear of flying. Results show experiencing a virtual airplane is just as effective as experiencing a real plane in reducing fear of flying anxiety. Both types of therapy are significantly better than no therapy at all. Others have looked at human perception of real and virtual stimuli. To explore the use of VR for lighting and color planning in buildings, Billger (2001) examined the perception of color in virtual and real environments. Wullemin et al. (2005) looked at differences in the perception of virtual and real spheres presented visually and with haptics. They found that virtual spheres presented visually are perceived as larger than real spheres of the same size. Slater et al. (2000) looked at the social behavior of small groups in real and virtual environments. Immersed participants (those experiencing the virtual world in a head-mounted display) were viewed as leaders by their peers (seated at monitors) in the virtual scenario but not in the real environment. Furthermore, group accord was higher in the real environment.

### **The Development of DIANA and VIC**

Through an interdisciplinary collaboration, we have created an interactive virtual clinical scenario of a virtual patient (VP) with acute abdominal pain. Abdominal pain is one of the most common ailments encountered by doctors. It is also a basic scenario in patient-doctor interaction and communication skills education. The doctor begins diagnosis by asking the patient a series of questions about the pain (history of present illness). At this stage, the doctor is trying to ascertain more information,

such as the pain's location and character, symptoms exhibited, family history, current medication, and aggravation if certain motions are performed. Sample questions include "What brought you into the clinic today?", "How long have you had the pain?", and "On a scale from 1 to 10, please rate the pain." The patient's responses will guide the doctor down different routes of questioning. The doctor evaluates the patient's response, gestures, and physical and auditory cues, such as wincing of pain, weight, posture, difficulty in making instructed motions, or pointing to specific areas. Based on asking the appropriate questions and evaluating the answers, treatment options can vary from immediate surgery to observation.

In the virtual scenario, a life-sized VP is projected on the wall of an exam room in a medical center. Before the virtual encounter, the student reviews patient information and receives directions include taking a history and developing a differential diagnosis. The virtual system includes two networked personal computers (PC's), one data projector, two cameras to track the users head and hand movement and a microphone. A commercially available speech recognition engine (Dragon Naturally Speaking Professional 8) was used to process the audio into phrases. The technology used in the study is readily available "off the shelf", and the entire prototype system cost less than \$7,000 (Figure 1).

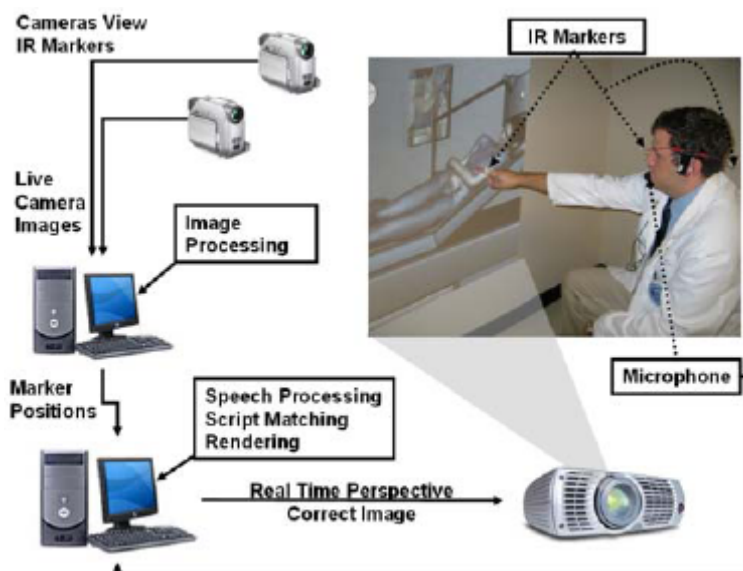


Figure 1

DIANA (DIgital ANimated Avatar), a female virtual character, plays the role of the patient with appendicitis, while VIC (Virtual Interactive Character), a male virtual character, plays the role of an observing expert (Figure 2). DIANA and VIC's scripts, which included gestures and audio responses, were created in consultation with several teaching medical faculty, with substantial standardized patient experience.



Figure 2

The student used speech and gestures to interact with DIANA and VIC. The system received audio and video input from the microphone and cameras. The audio was processed into phrases by the speech recognition engine. To improve accuracy, each participant created a voice profile. During the experience, the system displayed the recognized phrase on screen, allowing the student to identify if the system misrecognized a phrase. The history of present illness portion of the exam consists of a set of questions which the students are taught to ask. The script contained the most likely forms of each question, and several questions could map to the same response. For example, "Are you nauseous?" and "Have you been vomiting?" both result in DIANA telling the student that she has felt sick to her stomach. A simple established a cost to match the recognized phrase into each a question database, and then chose the lowest cost (below a threshold) as the understood question.

The system tracked the 3D trajectory of the students' hand with a marker-based tracking algorithm. Two gestures were recognized, handshaking and pointing. Handshaking was signaled if the student held their hand in front of their body for more than two seconds. Pointing was detected by finding the intersection of a ray (from the tracked head to the hand) and objects in the scene. A "laser pointer" red dot appeared where the system determined the student was pointing. While these were simple speech and gesture recognition techniques, they appeared adequate for the scenario.

Tracking the student's head position enabled DIANA and VIC's eyes to focus on the student. Correct perspective warping (Raskar, 2000) of the rendered image emphasized the characters' gaze directions, and maintained the illusion of the virtual examination room as an extension of the real room.

The system uses a simple state-based machine that transitioned between actions depending on input from the perception stage. Transition rules were based on accepted medical doctrine for the scenario. Actions included the virtual character speaking statements, changes in emotion, or animation. Our medical collaborators verified that acute abdominal pain diagnosis training lent itself well to this architecture.

DIANA and VIC are displayed at life-size using data projectors. This research proposes that seeing a human face and form at the appropriate size (as opposed to on a monitor) increases immersion and triggers psychological responses. The system used Haptik Inc.'s character animation library, which can generate high-quality, dynamic facial expressions and gestures. Secondary devices provided the student with information and more realistically simulated the encounter. The student used a TabletPC as a notepad and to receive scenario information (a mock 'patient file' is shown on the TabletPC at the beginning of the interaction. Afterwards, it is used as a notetaking device).

VIC's role in the experience is to welcome the student and instruct them on how to interface with the system (about a two minute tutorial). VIC then leaves the exam room, and the student proceeds to interview DIANA in a ten minute conversation.

We next present a recap of three studies that have been published and will analyze the studies and results from an education concepts perspective.

### **Study #1: Using virtual patients to teach communication skills (Stevens et al., 2005)**

An initial pilot study was conducted at the University of Florida with twenty participants. The purpose of the pilot study was to determine whether the virtual patient would be considered 'real' enough to use in later comparison studies with human, standardized patients. A prototype scenario of a patient with acute abdominal pain was directed at the second-year medical student level, recognizing that history-taking and communication skills are critical in the evaluation of a patient with abdominal pain.

After the exam, participants assessed the standardized patient by filling out the Maastricht Assessment of the Simulated Patient (MaSP; Wind, van Dalen, Muijtjens, & Rethans, 2004). The MaSP is a validated questionnaire that asks the medical student to rate the 'authenticity' of a standardized patient's portrayal of a condition. "The virtual patient stimulated me to ask questions" is an example MaSP question. Medical students who experienced DIANA also completed the MaSP questionnaire.

#### *Results.*

*Student Evaluation.* Students were surveyed using the MaSP following the exam to explore their evaluations about the tool and the technology behind the tool. The first part of the survey was on a 5-point Likert-type scale (1=Strongly Disagree, 5=Strongly Agree). On average, students believed that the tool appeared authentic ( $\mu = 3.95$ ) and stimulated them to ask questions ( $\mu = 3.75$ ). More importantly, they agreed that they would use the virtual scenario to practice their clinical skills ( $\mu = 4.25$ ). The second part of the survey assessed the students' beliefs about the technology; it was on a 7-point Likert-type scale (1=Least Important, 7=Most Important). Students reported a moderate level of sense of presence in the virtual exam room ( $\mu = 5.12$ ) and for suggesting that the VP gestures were life like ( $\mu = 5.67$ ). However, they found the most value in the fact that Diana was life-sized ( $\mu = 6.33$ ) and they wanted it to have a high quality of speech recognition ( $\mu = 6.71$ ).

*Study Summary.* In general, students were enthusiastic about the virtual interaction and its value as a teaching tool. In addition, their overall evaluation of the virtual scenario increased with subsequent versions as learner-centered suggestions for

improvement were incorporated. Most students felt the virtual interaction would aid in preparation for interaction with standardized and real patients. This study provided support for the notion that students were willing to interact with virtual patients and believed that they had a place in learning how to practice medicine.

*Implications for a Psychological & Pedagogical VR Framework.* Nass and Reeves' work (Reeves & Nass, 1996), which concentrates on what they term the 'media equation', offers evidence that humans enter into social contracts and relationships with technology. They argue that interactions with new media like television and computers are fundamentally social in nature. Much like interactions in real life, people expect media to obey a wide range of social and natural rules. Their research has provided VR developers with the understanding that given the right circumstances, humans will buy into the believability of an environment and act as they would with another human. However, that does not mean that any VR environment will work. In this environment, early speech recognition problems brought the students out of the relationship and made them cognizant of the product rather than the process. Improved recognition, although not perfect, allowed them to focus on the process of the interaction; as such, they valued the tool for its ability to help them practice their communication skills.

### **Study #2: An assessment of synthesized vs. recorded speech (Dickerson et al., 2005)**

In addition to testing the overall usability of the virtual patient system, it was important to evaluate specific features of the system that might hinder applicability. For instance, prior to comparing a virtual patient vs. human, standardized counterpart, one glaring difference is the voice of the human patient vs. the synthesized speech of DIANA. The purpose of this second study was to evaluate whether the type of speech made a difference in the use and usability of the system. If synthesized speech did not hinder the patient experience, its flexibility would enable a high level of interactivity. For example, DIANA could address each student by name and conversation changes would be easy to incorporate.

Seventeen medical students from the Medical College of Georgia participated in the study. All of the medical students were in their second or third year of study and each had several prior experiences with standardized patients. Participants were divided randomly into two groups with a system running with recorded speech ( $n=9$ ) or synthesized speech ( $n=8$ ).

There were three specific measures used to evaluate any possible differences between the two groups. First, a speech quality questionnaire for telephone dialogue systems (Möller, 2005) was adapted, targeting intelligibility, naturalness, pleasantness, comprehension, and overall acceptance of the voice. Sample questions include rating if "the voice was understandable". Second, the *Maastricht Assessment of the Simulated Patient* was used as in the first study. Finally, experts evaluated the tapes of the interactions and determined student task performance by identifying which core pieces of information, such as symptoms and signs, the student was able to elicit from DIANA including sections from chief complaint, history of present illness, and sexual history.

*Results.*



Learning objectives. No significant differences were found in the task performance ratings assigned by the experts between synthesized speech ( $\mu_{SS} = 4.37$ , s.d. = 1.59) and real speech ( $\mu_{RS} = 5.00$ , s.d. = 1.85). The ratings reflect the number of core questions asked during the interview. The SS condition presents lower fidelity audio than with RS, and may impact the effectiveness and believability of the simulation especially under more emotive scenarios. Synthesized speech allows the student to still meet educational objectives, and students scored DIANA equally under each condition for teaching ( $\mu_{RS} = 5.6$ , s.d. = 1.0,  $\mu_{SS} = 5.6$ , s.d. = 1.39,  $p = 0.46$ ) and training ( $\mu_{RS} = 5.1$ , s.d. = 1.12,  $\mu_{SS} = 5.1$ , s.d.=1.77,  $p = 0.49$ ).

Voice. Based on the questionnaire results, there was no reported difference in the intelligibility ( $\mu_{RS} = 4.9$ , s.d.=0.87,  $\mu_{SS} = 4.6$ , s.d. = 1.05,  $p = 0.28$ ), naturalness ( $\mu_{RS} = 4.3$ , s.d. = 0.65,  $\mu_{SS} = 4.2$ , s.d. = 1.22,  $p = 0.47$ ), and clarity ( $\mu_{RS} = 5.1$ , s.d. 0.82,  $\mu_{SS} = 5.0$ , s.d. = 1.75,  $p = 0.46$ ) of the voice. Some SS participants noted the synthetic speech sounded unnatural at first; however, they quickly stopped paying attention to the lack of prosody, and accepted the flow of conversation that the interface presented them. In the questionnaire, in reference to whether "this encounter is similar to other standardized patient encounters that I've experienced", there was some indication (but not a significant difference) that recorded speech is more familiar to students than synthesized ( $\mu_{RS} = 2.8$ , s.d. = 0.76,  $\mu_{SS} = 2.0$ , s.d. = 0.89,  $p = 0.05$ ).

The role of prosody. The role of prosody (non-verbal cues) is used to identify grammatical structure, convey attitude and emotion, and convey personal or social identity (Cohen, Giangola, & Balogh, 2004). However, the relative lack of prosody cues seemed to minimally impact this relatively simple scenario. The SS participants did not find SS limiting due to the simplicity of the VP's responses, the assumption that every response was a statement, and the simplicity of the conversation flow. Ambiguity did occur once in the scenario when the VP spontaneously asks the participant "can you help me!?" some SS participants were thrown off and had difficulty registering it as a question. Speech can show attitude and emotion, personality and social identity, however much of this information is visually presented. There may be a synergy of graphics and audio, and DIANA's expressive animation might have filled in what the audio had missing. Prosody appears more important for speech-only systems.

Study Summary. The results indicate no significant difference in performance between Group SS and Group RS in many of the task performance measures, such as the asking the correct questions. Upon closer inspection, there exist subtle – yet important – differences between virtual patients and standardized patients, primarily relating to conversation flow and the significant difference in level of expressiveness. Part of the lowered expressiveness is auditory, and thus SS's lower level of emotive expression impacts the overall experience. Recorded speech appears to be required to explore higher order communication skills. Our conclusions are as follows:

For lower level learning of communication skills, there appears to be little difference between RS and SS. Thus if the goal is to teach the student to recall which questions to ask, SS provides a compelling dynamic approach with minimal loss of educational objectives. However, if the goal is to teach the student how to ask the correct questions, (higher level learning) a high level of expressiveness in the virtual patient is needed. Essential information of the patient's condition could be lost from using synthesized speech. This in turn necessitates the higher cost – even with the lower flexibility – of recorded speech.

*Implications for a Psychological & Pedagogical VR Framework.* As with the first study, students were willing to buy into the believability of the VR tool, akin to the findings from Reeves & Nass (1996). However, it is important to understand the true nature of polymodal development. Multi-modal development means that there are multiple media sources present at the same time. Polymodal is term adapted from the biological sciences to refer to two or more media that appear at the same for the purpose of enhancing the other. In other words, they work together to support the overall goal of the VR tool. There are circumstances, particularly at the lower levels of Bloom's taxonomy, where virtual reality components such as speech can carry lower levels of fidelity. However, at higher levels, stronger fidelity must be attained if available and possible. Where fidelity is not truly achievable, it may be possible to supplement the tool with additional cues (i.e. visual) that support the overall educational goals of the environment.

### **Study #3: Comparing interpersonal scenarios**

Given the overall positive feedback on the system from the previous studies, a large controlled study was executed to compare students experiencing *either* DIANA or a standardized patient with the same symptoms (both working from the same script). This study (Raij et al., 2006) set out to examine the similarities and differences in experiencing an interpersonal scenario with real and virtual humans. Twenty-four medical students were assigned to one of two treatment groups. The standardized patient or *SP group*, which consisted of eight 2<sup>nd</sup>-year medical students from the University of Florida interviewed a real standardized patient named Maria. The virtual patient or *VP group* was from the Medical College of Georgia. Nine medical students and seven physician-assistant students interviewed the virtual patient.

At the conclusion of the experiment, medical experts from both institutions independently watched video recordings of both the real and virtual interactions. They assessed participants from Group SP and VP using behavioral measures, such as eye contact and appropriateness of conversation. The interactions were also analyzed for empathetic behavior. Empathetic behavior was judged by practicing clinicians as 'appropriate levels of empathy'. In practice, this usually involved a comment by the student to DIANA expressing her fears. Empathizing with the patient is an important skill that lets the patient know the doctor understands her situation (Coulehan & Block, 1997).

*Results.* The results from this study found similarities and differences between the virtual and real interpersonal scenarios in five key areas: Participant performance, participant behavior, scenario authenticity, patient expressiveness, and overall educational goals.

*Participant Performance.* Overall performance was similar between the two groups; both groups tended to elicit the same information from the patient and tended to ask the VP and SP the same questions. The medical expert reviewers agreed that at a high level, the interactions and task performance of Group VP and Group SP were similar. This supports the external validation of the virtual scenario as having a strong correlation to its real world counterpart. It also shows participants put the same effort into achieving the goals of a virtual interpersonal interaction as they would in a real one.

*Participant Behavior.* The number of times Group SP and VP expressed verbal empathy to the patient was similar ( $\mu_{SP} = 2.2$ , s.d. = 1.4,  $\mu_{VP} = 1.3$ , s.d. = 1.1,  $p = 0.44$ ). The main difference in empathetic behavior related to touch and style. Some participants touched the SP's leg or the exam bed and held it there for a moment. Conversely, the physical wall between the virtual and real exam room made it impossible for participants to touch the virtual patient. Group VP also had to adapt their conversational style to the limitations of the virtual patient. They asked questions in a more constrained manner and appeared to be less engaged.

*Scenario Authenticity.* Participant responses showed significant differences in whether the patient appeared authentic ( $\mu_{SP} = 5$ , s.d. = 0.0,  $\mu_{VP} = 3.8$ , s.d. = 0.58,  $p = 9E-6$ ), whether the encounter was similar to other standardized patient encounters they had experienced and whether the patient might be a real patient. Upon examining the debriefing comments, it became clear Group VP evaluated the 'humanness' of the virtual patient, whereas Group SP judged the accuracy of the standardized patient to a real patient. This result is similar to Usoh et al's (2000) conclusion that people apply different standards when assessing real and virtual environments on presence questionnaires (Usoh, Catena, Arman, & Slater, 2000). The indirect measures focused attention on individual aspects of the interaction. This allowed participants to specifically assess components, as opposed to deriving their own interpretations of overloaded terms such as "realism" and "natural". A battery of indirect measures that specifically addresses different experiences component, (e.g. specifically asking eye contact, audio fidelity, and speech recognition quality) will yield a clearer picture of authenticity.

*Patient Expressiveness.* The virtual and standardized patients were considered equivalent in displaying appropriate eye contact. The virtual patient was programmed to look at the participant during the interaction. This gaze behavior, life-size imagery, and rendering the exam room from the perspective of the participant contributed to sense that the virtual patient used appropriate eye contact. One Group VP participant commented: "I felt that it was neat that they were life-size, you know, and that the patient is looking at you and talking to you." However, the standardized patient expressed herself very differently from the virtual patient. Student feedback showed their beliefs that the SP communicated how she felt better than the VP ( $\mu_{SP} = 4.8$ , s.d. = 0.46,  $\mu_{VP} = 3.6$ , s.d. = 1.2,  $p = 0.005$ ) and a trend to be a better listener ( $\mu_{SP} = 4.5$ , s.d. = 0.53,  $\mu_{VP} = 3.5$ , s.d. = 1.2,  $p = 0.012$ ). The expressiveness of real people sets the bar very high for virtual characters. Participants specifically suggested that the VP be more expressive: "I would suggest to have more emotions into them. Maybe if there was more feelings, more emotional expression." Differences in performance may be a result of the virtual patient's poor expressive behavior. In general, the SP had more emotion in her voice (even compared to the same actress being the voice talent for DIANA), and her facial expressions and gestures were more 'believable' (for a lack of a better term).

*Overall Educational Goals.* The virtual and real scenarios were equivalent in student impressions of the educational value of the experience. Educational goals were clearly met by the virtual interaction despite the system's deficiencies.

*Study Summary.* Results of this study show the virtual patient was not nearly as expressive as the standardized patient. This contributed to differences in the conversational flow and less rapport with the virtual patient. However, the virtual interaction was found to be similar to the real interaction on many important education measures. Participants elicited the same information from both virtual and

standardized patients, and performed equally well overall. Furthermore, participants rated both interactions as equally valuable educational experiences.

*Implications for a Psychological & Pedagogical VR Framework.* Salomon & Gardner (1986) made the claim that educational research on computers could fall prey to the same mistakes and blunders of past research on the use of educational television. They specifically addressed the problem of asking questions that compare the effectiveness of learning in one medium vs. another. Swan (2003) called these studies “no significant difference” research, and demonstrated the point that Salomon and Gardner made that these questions are naïve and potentially useless. This research study essentially set out to examine student outcomes in one medium (human, standardized patient) vs. another (virtual patient). However, the goal of this project was not to prove the usefulness of one *instead of* the other. The purpose of this research, and in some sense the hope, was to demonstrate no significant difference to suggest that virtual patients could be a suitable alternative in the learning scenario. That does not mean that future research should continue to compare real vs. electronic because the scenarios and media are different. Instead, VR research should seek to explore why any differences might exist and what strengths could be brought from one medium to the other; or the goal could be to understand under what conditions the use of one might be of more educational value than the other. The goal should not be to prove one is better than the other in all circumstances. In addition, most VR research has concentrated on sight and sound. This research found that in some circumstances, touch was important to displaying empathy. Smell, touch, and perhaps even taste need to be explored to the extent that they meet the psychological and pedagogical goals of the learning environment.

### **Recommendations for future research**

Using important, research-based, pedagogic principles, we developed DIANA and VIC to help medical students learn communication skills. There are many benefits of this system. First, VIC can act as the scaffolding support that students need to learn complex skills. Research has provided evidence that the computer can be the more knowledgeable other in the student’s Zone of Proximal Development. In this case, VIC acts as the support mechanism. Scaffolding can then be slowly removed as the student becomes fully enculturated into the legitimate community of practice. Second, this VR system has demonstrated that it is possible to not only provide explicit ways for students to create artifacts (writing on the tablet PC), but also implicit ways that can be used to help them learn. For instance, this virtual reality system utilizes tracking devices that help students monitor where they are looking during an exam. Third, we have demonstrated the possibility of providing an environment where students can get repetitive practice on authentic, meaningful problems. This practice not only provides feedback, which is crucial to learning, but it also acts as a cost-effective and somewhat objective way to learn. A medical student could practice 40 or 50 times in a row at 3 or 4 in the morning with no real added cost to the medical college.

More importantly, the design, development, and implementation of DIANA and VIC has demonstrated both the possibility of designing pedagogically-sound virtual realities *as well as* evidence towards the production of new environments. Research in these studies suggested numerous important outcomes. First, we do not have to convince students that the virtual patient is real. In multiple trials, they were less concerned about the reality of the tool and more concerned with its ability to help them learn. In addition, almost all students believed that it was useful, it would help

them improve their skills, and they were willing to continue to work with the tool. This provides encouragement for the future development of such tools in multiple fields of education.

Second, we have provided evidence that at the lower levels of Bloom's taxonomy, virtual characters can lack certain features of expressiveness normally found in humans. However, for advanced tasks, the technology may not currently be proficient enough to rely solely on the visual expressiveness of the virtual character. Future development in this area or supplementation by other feedback cues might be necessary. In addition, future research should continue to explore a relationship between the taxonomy of the learning outcome and the necessary features of virtual character. Such an approach would prevent production and programming overkill and would promote investigations into the creation of integrated multi-modal environments. More specifically, we have provided evidence that there are some scenarios where synthesized speech is just as effective and at half the cost and flexibility of real voice.

Third, the effectiveness of VR tools in education is in some direct relationship to the feeling of presence a student gets while using the tool. However, that presence can be directly affected in both positive and negative ways by tools that may or may not be crucial to the environment. For instance, although perspective rendering is a useful concept, in this scenario it did not necessarily add to the outcome in relationship to the cost (or potential downside) of the feature. Conversely, spending more time on the script to achieve 90% recognition was a more useful objective that led to positive, observable outcomes. The same was true with using life-size characters. Our research knowledge needs to be strengthened related to various issues of presence and the cost/benefit ratio of each nuance of the VR system.

Finally, in a comparison of standardized patients and virtual patients, we have provided evidence of similar effectiveness as measured by student performance. This provides the most convincing evidence that designing correctly, in pedagogically-strong ways, VR tools such as virtual characters do hold a promising future for teaching and learning.

## References

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine*, 68(1), 52-81.

Badler, N.I., Erignac, C.A., & Liu, Y. (2002). Virtual humans for validating maintenance procedures. *Communications of the ACM*, 45, 56 – 63.

Bailenson, J.N., Aharoni, E., Beall, A.C., Guadagno, R.E., Dimov, A., & Blascovich, J. (2004). Comparing behavioral and self-report measures of embodied agents: Social presence in immersive virtual environments. Paper presented at *Proceedings of the 7th Annual International Workshop on PRESENCE*, Valencia, Spain.

Bailenson, J.N., Blascovich, J., Beall, A.C., & Loomis, J.M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 10, 583-598.

Batson, T. (1993). The origins of ENFI. In B. C. Bruce, J. K. Peyton, & T. Batson (Eds.), *Network-based classrooms: Promises and realities* (pp. 87-112): Cambridge University Press.

Billger, M. (2001). Colour appearance in virtual reality: A comparison between a full-scale room and a virtual reality simulation. Paper presented at *The 9th Congress of the International Colour Association*.

Blumenfeld, P. C., Krajcik, J. S., Marx, R. W., & Soloway, E. (1994). Lessons learned: How collaboration helped middle grade science teachers learn project-based instruction. *The Elementary School Journal*, 94(5), 539-552.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.

Brown, A. L., & Ferrara, R. A. (1985). Diagnosing zones of proximal development. In J. V. Wertsch (Ed.), *Culture communication, and cognition: Vygotskian perspectives* (pp. x, 379). Cambridge [Cambridgeshire]; New York: Cambridge University Press.

Cohen, M., Giangola, J. & Balogh, J. (2004). *Voice user interface design*. Addison-Wesley.

Cole, M. (1996). *Cultural psychology : A once and future discipline*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Coulehan, J.L. & Block, M.R. (1997). *The medical interview: Mastering skills for clinical practice* (3<sup>rd</sup> ed.). Philadelphia: F.A. Davis Company.

Dickerson, R., Johnsen, K., Raij, A., Lok, B., Bernard, T., Stevens, A., Lind, D.S. (2005). Virtual patients: Assessment of synthesized versus recorded speech. To appear in the *Proceedings of Medicine Meets Virtual Reality (14)*.

Emmelkamp, P.M.G., Krijn, M., Hulsbosch, A.M., de Vries, S., Schuemie, M.J., & van der Mast, C.A.P.G. (2002). Virtual reality treatment versus exposure in vivo: A comparative evaluation in acrophobia. *Behaviour Research and Therapy*, 40, 509-516.

Engestrom, Y. (1991). Activity theory and individual and social transformation. *Multi-disciplinary Newsletter for Activity Theory*, 7/8, 6-17.

Ferdig, R.E. (2005). Towards implementing technologies in education: Exploring the pedagogy and people of good innovations. *Turkish Online Journal of Educational Technology*, 4(2), [Online Serial]; 17 manuscript pages.

Garau, M. Slater, M., Bee, S., & Sasse, M.A. (2001). The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. Seattle, Washington, United States: ACM Press, 2001.

Gavelek, J. R., & Raphael, T. (1996). Changing talk about text: New roles for teachers and students. *Language Arts*, 73, 182-192.

Harré, R. (1984). *Personal being: A theory for individual psychology*. Cambridge, M.A.: Harvard University Press.

Harré, R., Clarke, D., & DeCarlo, N. (1985). *Motives and mechanisms*. New York: Methuen.

Hill Jr., R.H., Gratch, J., Marsella, S., Rickel, J., Swartout, W., & Traum, D. (2003). Virtual humans in the mission rehearsal exercise system. *Kynstliche Intelligenz (KI Journal)*, 17, 2003.

Johnsen, K., Dickerson, R., Raij, A., Lok, B., Jackson, J., Shin, M., Hernandez, J., Stevens, A., Lind, D.S. (2005). Experiences in using immersive virtual characters to educate medical communication skills. In *Proceedings of IEEE Virtual Reality 2005*.

Kolodner, J. L. (1997). Educational implications of analogy. *American Psychologist*, 52(1), 57-66.

Krajcik, J. S., Blumenfeld, P.C., Marx, R.W., Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The Elementary School Journal*, 94(5), 483-497.

Lang, F., McCord, R., Harvill, L., & Anderson, D. (2004). Communication assessment using the common ground instrument: Psychometric properties. *Family Medicine*, 36(3).

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Linn, M. C. (1991). The computer as learning partner: Can computer tools teach science? In K. Sheingold, Roberts, Linda G., and Malcom, Shirley M. (Ed.), *Technology for Teaching and Learning*. Washington D.C.: American Association for the Advancement of Science.

Manganas, A., Tsiknakis, M., Leisch, E., Ponder, M., Molet, T., Herbelin, B., Magnenat-Thalmann, N., Thalmann, D., Fato, M., & Schenone, A. (2004). The Just Vr Tool: An innovative approach to training personnel for emergency situations using

virtual reality techniques. *The Journal on Information Technology in Healthcare*, 2, 399-412.

Möller, S. (2005). *Quality of telephone-based spoken dialogue systems*. Springer: New York, NY.

Mortensen, J., Vinayagamoorthy, V., Slater, M., Steed, A., Lok, B., & Whitton, M. (2002). Collaboration in tele-immersive environments. In *Proceedings of the Eighth Eurographics Workshop on Virtual Environments*, May 30-31, 2002.

Newman, F. M., Secada, W. G., Wehlage, G. (1995). *A guide to authentic instruction and assessment: Vision, standards and scoring*. Madison, Wisconsin: Wisconsin Center for Education Research.

O'Shea, T. (1999). *Birbeck Web forum on learning and teaching*. Available: <http://www.bbk.ac.uk/asd/view/view02.html>.

Oviatt, S.L., Cohen, P.R., Wang, M., & Gaston, J. (1993). A simulation-based research strategy for designing complex NI systems. *Paper presented at ARPA Human Language Technology Workshop*, Princeton, N.J.

Pertaub, D., Slater, M., & Barker, C. (2001). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, 11(1), 68-78.

Raij, A., Johnsen, K., Dickerson, R., Lok, B., Cohen, M., Stevens, A., Bernard, T., Oxendine, C., Wagner, P., & Lind, D.S. (2005). Interpersonal scenarios: Virtual = real? To Appear in the *Proceedings of VR2006*.

Raskar, R. (2000). Immersive planar displays using roughly aligned projectors. In *Proceedings of IEEE Virtual Reality 2000*.

Reeves, B., & Nass, C. I. (1996). *The media equation : how people treat computers, television, and new media like real people and places*. Stanford, Calif. New York: CSLI Publications ; Cambridge University Press.

Roehler, L., & Cantlon, D. (1996). *Scaffolding: A powerful tool in social constructivist classrooms*. Available: <http://www.educ.msu.edu/units/literacy/paperlr2.htm>.

Rogoff, B. (1994). Developing understanding of the idea of communities of learners. *Mind, Culture, and Activity*, 1(4), 209-229.

Rothbaum, B.O., Hodges, L., Smith, S., Lee, J.H, & Price, L. (2000). A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, 68, 1020-1026.

Salomon G. & Gardner, H. (1986). The computer as educator: Lessons from television research. *Educational Researcher*, 15 (1), 13-19.

Sandford, S., & Richardson, K. (1997). Interactive instructional design: Old paradigms for new technologies. Paper presented at the *NAU/web.97*, Flagstaff, AZ.



- Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences*, 1(1), 37-68.
- Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators and Virtual Environments*, 9, 37-51.
- Stevens, Amy, Jonathan Hernandez, Kyle Johnsen, Robert Dickerson, Andrew Raij, Jonathan Jackson, Min Shin, Juan Cendan, Margaret Duerson, Benjamin Lok, D. Scott Lind (2005). The Use of Virtual Patients to Teach Medical Students Communication Skills. To appear in *American Journal of Surgery*.
- Swan, K. (2003). Learning effectiveness: what the research tells us. In J. Bourne & J. C. Moore (Eds) *Elements of Quality Online Education, Practice and Direction*. Needham, MA: Sloan Center for Online Education, 13-45.
- Thórisson, K.R. (1997). Gandalf: An embodied humanoid capable of real-time multimodal dialogue with people. Paper presented at the *First ACM International Conference on Autonomous Agents*, Marina del Rey, California.
- Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9 (5), 497-503.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, Mass.: Harvard University Press.
- Wind, L.A., van Dalen, J., Muijtjens, A.M.M., & Rethans, J.J. (2004). Assessing simulated patients in an educational setting: The Masp (Maastricht Assessment of Simulated Patients). *Medical Education*, 38, 39-44.
- Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring and problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89-100.
- Wuillemin, D., van Doorn, G., Richardson, B., & Symmons, M. (2005). Haptic and visual size judgements in virtual and real environments. Paper presented at the *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'05)*.