

2011

The Effectiveness Of Virtual Humans Vs. Pre-recorded Humans In A Standardized Patient Performance Assessment

Joel Palathinkal
University of Central Florida

 Part of the [Psychology Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Palathinkal, Joel, "The Effectiveness Of Virtual Humans Vs. Pre-recorded Humans In A Standardized Patient Performance Assessment" (2011). *Electronic Theses and Dissertations, 2004-2019*. 1788.
<https://stars.library.ucf.edu/etd/1788>

THE EFFECTIVENESS OF VIRTUAL HUMANS VS. PRE-RECORDED HUMANS
IN A STANDARDIZED PATIENT PERFORMANCE ASSESSMENT

by

JOEL PALATHINKAL

B.S. University of South Florida, Electrical Engineering 2004

M.S. Iowa State University, Systems Engineering 2006

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Modeling and Simulation
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2011

Major Professor: Peter Kincaid

© 2011 Joel Palathinkal

ABSTRACT

A Standardized Patient (SP) is a trained actor who portrays a particular illness to provide training to medical students and professionals. SPs primarily use written scripts and additional paper-based training for preparation of practical and board exams. Many institutions use various methods for training such as hiring preceptors for reenactment of scenarios, viewing archived videos, and computer based training. Currently, the training that is available can be enhanced to improve the level of quality of standardized patients. The following research is examining current processes in standardized patient training and investigating new methods for clinical skills education in SPs. The modality that is selected for training can possibly affect the performance of the actual SP case.

This paper explains the results of a study that investigates if there is a difference in the results of an SP performance assessment. This difference can be seen when comparing a virtual human modality to that of a pre-recorded human modality for standardized patient training. The sample population navigates through an interactive computer based training module which provides informational content on what the roles of an SP are, training objectives, a practice session, and an interactive performance assessment with a simulated Virtual Human medical student. Half of the subjects interact with an animated virtual human medical student while the other half interacts with a pre-recorded human. The interactions from this assessment are audio-recorded, transcribed, and then graded to see how the two modalities compare. If the performance when using virtual humans for standardized patients is equal to or superior to pre-recorded humans, this can be utilized as a part task trainer that brings standardized patients to a higher level of effectiveness and standardization. In addition, if executed properly, this tool could potentially be used as a part task trainer which could provide savings in training time, resources, budget, and staff to military and civilian healthcare facilities.

This dissertation is dedicated to the people who continue to push technology and innovation despite the discouragement given from skeptics. Scientific advancements, when not discovered by mistake, were created through perseverance and faith. When not solely through perseverance, individuals of the past who have stood through adversity to take a stand for what they believe in; have motivated the individuals of present day. In the future, we should continue making these bold strides; as each influential milestone will be a stepping stone for mankind, progress, and peace on earth.

-Joel Palathinkal

“For someone who was never meant for this world, I must confess I’m suddenly having a hard time leaving it. Of course, they say that every atom in our bodies was once part of a star. Maybe I’m not leaving....maybe I’m going home.”

-GATTACA

ACKNOWLEDGMENTS

It is impossible to quantify the actual time and effort it takes to raise a child. Not only does it take two loving parents, but it truly takes a village to raise a child. My parents (John and Elsamma Palathinkal) utilized the values and beliefs that were instilled upon them to constantly nurture a strong marriage; which later fortunately became the backbone of a strong family unit. My brother Joby constantly kept me laughing. These aspects later reflected the way I was raised; and allowed me to be grateful for all the blessings that I was given. I am forever gracious for having a supportive family.

Bill Clinton once said that his major successes were significantly relevant to the friendships that were developed in his lifetime. In addition, Woody Allen said that 80% of success is showing up on time. I completely agree with both of these strong points as if I didn't have the strong professional, personal, and educational support groups; along with the perfectly timed instances of opportunities, I would not be where I am today.

A special and important thanks goes to all that were involved in the development of this dissertation. These members consisted of all the faculty and supporting personnel at the University of Central Florida's College of Medicine, College of Nursing, Institute for Simulation and Training, Florida Hospital's Nicholson Center for Surgical Advancement, the Army Research Development and Engineering Command, and the University of Florida's Virtual Environments Research Group (VERG) team which is part of the Computer and Information and Science and Engineering Department. The VERG team provided crucial technical insight that was needed to allow me to use the SP case, Roleplay trainer, and have a better understanding of conversational modeling. A special gratitude goes out to Dr. Moshe Feldman for mentorship in medical education along with Dr. Richard Satava, who has and still continues to grace me with his expertise, valuable experience, and forward thinking strategy for the future of healthcare. Furthermore, I would like to especially acknowledge my

colleague and mentor Curtis Conkey as he has opened doors for me from the networking of industry leaders, to introductions to recognized technologists. In addition, a sincere thanks goes to my mentors Maynard Zettler, Dan Dwyer, Kenneth (Chip) Reams, CAPT Roto Reuter, Michael Merritt, and Walt Augustin at the Naval Air Warfare Center Training Systems Division (NAWCTSD), the Acquisition Workforce Tuition Assistance program of the United States Navy, Dr Sheehy for approving funding to continue my dissertation, and the Department of Defense who provided critical support to this PhD effort. Without their logistical and financial support, this dissertation would have been very difficult to achieve. In addition, several employees at the NAWCTSD and students at UCF served as research subjects for my PhD study. Most importantly, I would like to thank my very supportive PhD committee (Dr. Randy Shumaker, Dr. Roger Smith, and Dr. Kelly Allred) and my dissertation chair advisor Dr. Peter Kincaid for having faith and support in my pursuits.

TABLE OF CONTENTS

LIST OF FIGURES	xi
LIST OF TABLES	xiii
CHAPTER ONE: INTRODUCTION	1
Statement of the Problem	1
Purpose of the Current Study	6
CHAPTER TWO: LITERATURE SEARCH.....	7
How Are Standardized Patients Trained	7
Coaching Standardized Patients	7
Computer Based vs Traditional training	9
Crossover study between Standardized Patient and Web based training	13
Medical Student Performance with rich media	17
Rater Training.....	20
Rater Rapport	21
Interrater Reliability	22
Bias and Poor Interrater Reliability	22
Causes for Bias and Lack in Reliability	23
Enriched rater training using Internet based technologies	24
Avatar Immersion and Presence	27
Facial Expressions in Communication	28
Designing Embodied Conversational Agents.....	29
Embodied Conversational Agent Allies	32
Team Based Training	34
Communicative Tasks	39
Embodied Conversational Agents and Simulated Personalities.....	39

ECAs that Present and Sell.....	42
Smart Avatars	44
Embodied Characters	45
Evaluating Interactive Systems with Embodied Conversational Agents	47
Ethnicities in Embodied Conversational Agents.....	47
Human Centered Distributed Conversational Modeling	48
Bootstrapping the Creation of Virtual Humans from existing ones for SP Training	49
Participant Motives in Conversational Modeling.....	52
Virtual Patient Interactions.....	53
Virtual Humans Providing Mentorship	55
Mental Models and Immersion of the Future	56
Trusting Virtual Humans and Synthetic Thinking	57
Acting Concepts	57
Experiential Learning	58
Medical Focus Group	58
Medical Student Feedback	60
Research Statement	62
CHAPTER THREE: RESEARCH METHOD	64
Scope	64
Test	64
Hypothesis	66
Manipulation	66
Summary of Manipulation.....	67
Power Analysis.....	67
Recruitment	68

Study Workflow and Timeline	68
The timeline for the study	70
Required Equipment and Supplies:	70
Objective/Subjective Measurements:	71
Decision Rules.....	71
Justification for Exclusion of Specific Groups:	71
Risks and Discomforts to Research Volunteers	72
Maintenance of Experimental Data for Records.	72
CHAPTER FOUR: RESULTS	74
Demographics.....	74
Demographics Tables	75
Open Ended Feedback Results	76
Hypothesis Testing	76
CHAPTER FIVE: DISCUSSION.....	80
Natural Language Processing	80
Future Approaches	80
Coding for Interrater Reliability	80
Using Real SPs.....	80
Automated Assessments	81
Leveraging the Acting Community	81
Sixth Sense Technology.....	82
Holograms.....	83
Stretching Visual Images to Reduce Pain	84
Similarity Effects to Teaching	84
Using Artificial Intelligence for Resolving Disputes.....	85

Publication Status	85
Conclusion.....	85
Discoveries	86
APPENDIX A: IRB APPROVAL FORMS	88
APPENDIX B: STUDY PACKET	92
APPENDIX C: SLIDES FOR INTRODUCTION AND PRACTICE SESSION	107
APPENDIX D: SLIDES FOR VIRTUAL HUMAN PERFORMANCE ASSESSMENT...	113
APPENDIX E: SLIDES FOR PRE-RECORDED HUMAN PERFORMANCE ASSESSMENT	117
APPENDIX F: ADDITIONAL DATA/ARTIFACTS.....	121
REFERENCES	128

LIST OF FIGURES

Figure 1: Video/Avatar Screenshot Comparison	5
Figure 2: Study Flow Diagram	15
Figure 3: Rea Model	30
Figure 4: Speech and Gesture generation with SPUD server	31
Figure 5: Three layers: scheduling, animation, and rendering.....	32
Figure 6: Synchronized speech with the action scheduler	32
Figure 7: Issues in design for ECAs	33
Figure 8: Top Level ECA Architecture	33
Figure 9: FXPAL ECA architecture	34
Figure 10: Training to Extinguish an alarm.....	34
Figure 11: Cognition, Perception, and Motor Control.....	35
Figure 12: Deictic test bed with Cosmo the interactive agent	37
Figure 13: Agent behavior planning architecture	38
Figure 14: Anatomical Emotive Information.....	38
Figure 15: Expressions- Suggesting, imploring, and ordering	39
Figure 16: Position of emotions within Valence-Arousal space.....	40
Figure 17: Position of personality within Dominance-Friendliness space	41
Figure 18: Agents used in sales	43
Figure 19: Assuming roles for agents in sales	43
Figure 20: Jack's Moose Lodge	44
Figure 21: VET environment.....	45
Figure 22: Angry Facial Expression in VET	45
Figure 23: Polygon wireframe for an ECA.....	46
Figure 24: Syllable Pronunciations.....	46

Figure 25: HCDCM Design	48
Figure 26: Virtual People Factory Workflow	49
Figure 27: Virtual Medical Student Generation.....	50
Figure 28: Comparing Preparedness between the Pre and Post Survey	51
Figure 29: Comparing Confidence between the Pre and Post Survey	51
Figure 30: Leaderboard for Discoveries	53
Figure 31: Narrative	53
Figure 32: Life-size Bipolar Virtual Patient	54
Figure 33: Web based Virtual Patient	54
Figure 34: Five Levels of Hierarchy	55
Figure 35: Study Design	66
Figure 36: Sixth Sense Technology	82
Figure 37: Sixth Sense Virtual Tablet PC.....	83
Figure 38: Sixth Sense Technology: Projection of Feeds	83
Figure 39: Standardized Patient Questionnaire 1.....	126
Figure 40: Standardized Patient Questionnaire 2.....	127

LIST OF TABLES

Table 1: Training Modality Comparison Matrix	12
Table 2: Modality, Gender, Race, and Age Demographics	75
Table 3: Education, Acting, and Healthcare Demographics	75
Table 4: Kolmogorov-Smirnov test of Normality	77
Table 5: Training Method Homogeneity	77
Table 6: Results of a T-Test using Two Sample Assuming Unequal Variances	78
Table 7: Virtual Human vs Pre-Recorded words per minute comparison	87
Table 8: Additional Tukey Analysis	122
Table 9: Pre and Post Survey PANAS scales	123
Table 10: Perceived Usefulness	124
Table 11: Perceived Ease of Use	125

LIST OF ACRONYMS

AI.....	Artificial Intelligence
CEX.....	Clinical Evaluation Exercise
DARPA.....	Defense Advanced Research Project Agency
ECA.....	Embodied Conversational Agent
HCDCM.....	Human-Centered Distributed Conversational Modeling
IITSEC.....	Interservice/Interindustry Training, Simulation, and Education Conference
IRB.....	Institutional Review Board
IPS.....	Interpersonal Simulator
Cex.....	Clinical Evaluation Exercise
NAWCTSD.....	Naval Air Warfare Center Training Systems Division
NEX.....	Neurology Evaluation Exercise
NLP.....	Natural Language Processing
OSCE.....	Observed Structured and Clinical Evaluations
RFI.....	Request for Information
RITE.....	Residency In-service Training Examination
SOAP.....	Subjective, Objective, Assessment, and Plan
SME.....	Subject Matter Expert
SP.....	Standardized Patient
UI.....	Urinary Incontinence
VERG.....	Virtual Environments Research Group
VH.....	Virtual Human
VMS.....	Virtual Medical Student
V-people.....	Virtual People
VPF.....	Virtual People Factory
WB.....	Web Based
3-D.....	Three Dimensional

CHAPTER ONE: INTRODUCTION

Statement of the Problem

Effective and accurate training is needed to enable Standardized patients (SPs) to be concurrent with the medical technology and training objectives that are constantly changing. The training that is currently available has not evolved much, and this evolution is needed for future growth and sustainment of training that is heavily reliant on advanced simulation and training. The use of Standardized Patients in role-play training to learn clinical skills and medical interviews has been existent since 1963 (Rossen, 2010). A typical standardized patient interaction includes a role playing scenario that lasts five to thirty minutes. When this interaction occurs, questions are asked from the physician or medical students based on the medical case that is learned from the SP (Rossen, 2010). The medical student or doctor interacts with the patient and gives them a diagnosis based on the physical and verbal cues that a patient provides (Rossen, 2010). Standardized patients are actors who are hired to portray a patient who possesses a particular medical condition (Rossen, 2010). These SPs perform this same type of roleplaying task with several different students and must act consistently the same to maintain the integrity of the curriculum and standardized board exam (Rossen, 2010). Attaining this skill to be able to successfully resemble the same patient and scenario with each student is essential to providing a standardized training solution especially when evaluating students. This role-play interaction, aside from training, is also used as an evaluation metric for the Step 2cs (clinical skills exam). When interviewing a few medical students, the conclusion that was gathered is that the result of passing is not as significant as to failing the step 2cs. If the board exam is failed, it is be very challenging to enter into residency program. Another issue is that some standardized patients lead the

student on with answers that could suggest an unfair advantage to some students. In order to have a good test, the SPs must be standardized

As mentioned in the Abstract, a major responsibility of a standardized patient is to be able to portray a patient with the fluent knowledge of the symptoms and also be able to properly rate the medical student after the interaction by accurately completing a Standardized Patient checklist. Having the standardized patient effectively reenact the scenario in an identical standardized manner can be a challenge.

When comparing the actual verbal content to the accuracy, there is a difference. If you take the verbal content responses, they are correct within a 90.2% of correctness, while the accuracy comes within a 30% range of accuracy (Tamblyn, 2009). Some areas where some of the standardized patients can improve are: remembering how to respond correctly when a medical student asks questions, understanding what is asked, and being able to interact effectively and consistently (Rossen, 2010). There are moments, where the questions or combinations of questions could be unexpected or not able to comprehend (Rossen, 2010). The diversity of a medical student can create an area for error as well, as some individuals may have different accents in language, diction, and delivery. The VERG (Virtual Environments Research Group) at the Computer Information Science Engineering Department at the University of Florida has developed a Roleplay trainer which trains SPs to utilize the voice logs of previous medical students to assist in clinical training (Rossen, 2010).

The tool that is created can provide the standardized patients with a broader array of questions that could be asked by a medical student, preceptor, or physician. The training lies in being able to provide the correct responses to the particular questions that were asked, providing an appropriate response, and giving a form of feedback to answer what is asked. A more optimized solution in simulating the conversation that is delivered

by the Virtual Human, allows the SP to have a better level of proficiency in a standardized patient interaction (Rossen, 2010, Wessell, 2003).

According to Fussell, a community based treatment organization uses Standardized Patients to portray a substance abuse client. In order to reach a certain level of expertise, SPs in this scenario trained for 15 to 25 hours (Fussell, 2009). A good amount of this time consists of the tasks regarding memorization and periodic assessments of the material that is presented (Fussell, 2009).

Another instance of SP training is when another educator is used to assist in training SPs on a psychological case. This case interfaces with a virtual human that represents a realistic representation of physical and psychological symptoms (Fussell, 2009). The areas that are required for the SP to perform are created and built into the learning tasks of the training system. This system is good at being able to pinpoint these areas that could potentially be an obstacle for reaching a particular proficiency (Fussell, 2009). A focus group is used to evaluate the validity in portrayal of the cases. Experts in the areas of these learning applications are asked to observe these experiences, and subjects give a particular score on a checklist for each SP with the use of a Likert-Scale. The questions are based on different types of factors on the interactive experience such as how realistic it is (Fussell, 2009).

Multimedia is widely used as a method of training. Many times when learning modules are examined, the scope of the tool needs to be considered when trying to understand if it is used for assessment or training. After assessing the objectives and goals of the experiment, the learning module that is developed for this experiment can be considered a training tool. Part task training has been a popular form of training for critically focused skills to reduce cost and training time that is used in a more expensive multi-functional trainer. This poses true especially when the task is very specialized. An

example of this is seen in standardized patient training where SPs can practice in a virtual and immersive environment with a standardized medical student to help learn their lines, develop a better proficiency of the case, and also remove some instances of anxiety that may be linked to an actual human interaction with the help of this virtual rehearsal. This allows the training to occur on demand without having to hire complete faculty and additional resources to get the focused training needed on a particular skill set. This training for readiness does not replace a standardized patient training curriculum, but provides a positive supplement which provides an overall cost savings over a long period of time.

While there are varying definitions of multimedia, it can be generally defined as extending training beyond pure text using a combination of several potential media types such as graphics, animations, video, auditory, and photographs (Conkey, 2010). Indeed, technological advances continue to produce new media types that can be used in multimedia based training and educational venues (Conkey, 2010). In this case, the experiment uses an integrated multimedia training tool for Standardized Patient training.

A question that is often asked in an avatar vs. video experiment is if avatars provide similar or exceeding results in training when compared to video delivery modalities. Avatars may not be at the same level of training fidelity that the video provides but the technology has improved to provide a better quality of realism so that many of the gestures, inflection, tone, eye movement, and emotion can be represented with simulation software, algorithms, decision modules, and artificial intelligence. A definition of Fidelity is how accurate a system is able to recreate an output which is identical to its input (Wordnet, 2010). This electronic system should assist the user in achieving a mastery of skills that are needed to complete a task on the job. Training fidelity is essential in providing a robust training solution that provides the user an

experience that is immersive and also provides retention of a learning task. As seen in the screenshots below, the avatar to the right may not provide an exact replica of the character to the left, but what if the same training value is there?



Figure 1: Video/Avatar Screenshot Comparison

In addition, since the avatar is used, synthetic voice matching technology could be used to modify the voice track and alter it to accommodate new learning tasks. This could provide extensive savings as it takes a great amount of time, cost, and collaboration to produce videos with human subjects. Cost savings could also be incurred when edits to the voice sound file could be changed along with lip syncing to match so that the avatar could be reused to teach new content as opposed to all the rework that is required when editing video footage of human actors for a brand new training scenario. However, a recurring question is if avatars react in the same amount of credibility as that of the video actors. Little, if any, research has so far compared avatar actors with human actors for their effectiveness in standardized patient training. There are many dimensions of human interaction and perceptions that can affect the production of effective avatar based training (Conkey, 2010).

Conkey asks the following interesting questions: Will the trainee have the ability to have an immersive experience (level of presence) while using avatar actors versus human actors? Furthermore, if the presence values are negatively affected, how will the training outcomes be affected?

Purpose of the Current Study

This study analyzes the results of a performance assessment of two distinct modalities. The findings will evaluate measurements on if there is an enhancement in performance with standardized patients who were trained with Virtual Humans in comparison with the Pre-recorded human. These two learning objectives will be analyzed upon the performance assessment.

CHAPTER TWO: LITERATURE SEARCH

How Are Standardized Patients Trained

There are various methods used currently to train standardized patients. Some training approaches utilize integrated methods, while others are more focused. The following sections below will provide details on different methods and comparisons on how SPs are trained. In addition, explorations will be made when comparing different modalities and if there are proven results or no change when considering interactive training.

Coaching Standardized Patients

One method of training standardized patients is coaching. Coaching can produce skills that are enhanced in the level of effectiveness, and this depends on how well the coach leads their cohort of people with the proper guidance (Wallace, 2007). In order for a case to be more valid, the SPs should be at a similar level of performance so that some people are not getting better training than others. A selection of quality standardized patients will allow an SP case to be represented in a more accurate and realistic sense (Wallace, 2007). There is a quite involved collaboration in the Standardized patient training arena when it comes to creating a well written clinical case, allowing the SP to bring the patient to life, and continually improving coaching abilities (Wallace, 2007).

While developing coaching abilities, there are some qualities that Peggy Wallace, the author of “Coaching of Standardized Patients” refers to as qualities of an effective coach. One of the first traits that are required is knowing something about acting and directing (Wallace, 2007). Since the standardized patient experience is highly involved with patient portrayal, the coach must know something about acting and directing to be able to properly train and simulate an actual standardized patient encounter. A second

trait is being able to develop a trusting relationship with SPs (Wallace, 2007). As quoted in Funke & Booth, 1961, p. 285, the relationship should be one of mutual respect, understanding, having love of the work, and also having trust. Because the director (coach) represents the eye which helps the actor (SP) to decide the form in which he does takes place; it is the coach's vision of the patient case that enables the SPs to develop the "form" in which they portray the patient (Wallace, 2007). A third trait is bringing enthusiasm and sensitivity to their work (Wallace, 2007). Bringing this imagination in the SPs mind is one of the key factors that represent patient portrayals. When a coach successfully demonstrates this skill, the SP is motivated to bring the patient case to life in its true essence. Finally, the fourth trait is to trust their intuition and what they know. This intuition must come from within, and they must use that to help guide them in their interactions. As SPs use their instincts during patient scenarios, the actual coaches also use an instinctual approach to respectfully learn what is possible and then proceed with current knowledge (Wallace, 2007). Being able to duplicate these characteristics in training provides the ability to produce a quality curriculum.

To become an SP, the SP must have the ability to portray a patient, observe the way that the medical student acts, remember the encounter, complete the checklist based on what is seen, and provide insightful feedback to the student so they may be able to debrief the scenario (Wallace, 2007). As mentioned before, the training on standardized patients is used to perform in any high-stakes clinical skills assessments. The preparation of this requires that SPs have seamless portrayals that are very similar to how an actual patient would behave. In addition, the SP would be able to provide challenges for the medical student examinees, and have a flawless knowledge of the case (Wallace, 151).

Wallace also suggests that there are particular guidelines that must be followed when considering taking care of SPs. One of the first steps is to keep in mind the specific

skillsets that are needed for each SP to attain (Wallace, 2007). Another is to ensure that all the SPs understand what is meant by the word standardization. Standardization does not always mean that all the SPs need to act in the same manner, but standardization is better described as having the SP portray a patient consistently in his/her own rendition relevant to the scenario. Furthermore, standardization may mean something different to SPs who have acting experience. Actors who are training may have a perception that it may be a negative trait to make a scene standardized. There are some responses and gestures that are situational. For example, a medical student's inflection in speech could provide more expectations to provide gestures that supplement the words that are put out. Standardization can also be seen as an overarching interpretation. An interpretation of standardization that is commonly used is making training available to a large group of people. This clear layout of a vision of the case reduces ambiguity. In addition, the issues, concerns, and examples that come up in training are the very substance of this single-minded understanding of the case among all SPs (Wallace, 2007).

According to Win May, an essential aspect in clinical skills training is to have the SP be able to evaluate medical students at an above average standard. This means that when they rate the medical students, the SPs are giving an answer that is correctly graded. These rating grades must be attained at an accuracy of at least 85% or above (May, 2008). To ensure that this occurs, the efforts of the SPs can be observed by subject matter experts (May, 2008).

Computer Based vs Traditional training

The standardized patient program at USC allows the patient to train to simulate the signs and symptoms of actual patients. Some good candidates for these are people who have stable, abnormal physical findings. The students that are chosen are able to

successfully reenact a scenario the same time, every time. This provides standardization in performance. (Standardized, 2009).

SPs don't use this method at the moment as this is normally used for the training of medical students. SPs were used in a study that lasted over the span of two years. This SP curriculum was combined with two methods. This program was broken out into a computer based and traditional treatment in training. Post-test counseling was the main training objective (Konkle, 2002). There were approximately 500 people or more of these individuals who were located in rural areas. These geographic areas provided positive results. The awareness and capability to perform in clinical practice was strengthened as a result. Furthermore, there is a strengthened sense of being able to mitigate and detect risks for the subjects.

The electronic delivery of training content for medicine is becoming a growing trend. Some of these computer based training tools have other features such as allowing a student to go back and remediate themselves when questions were answered incorrectly. However, there has not been significant research that discovers how these self-delivered educational programs provide a benefit in learning transfer for standardized patients (Errichetti, 2006). This study is essential in providing a comparison between the traditional techniques of acting out a scenario and being given training as opposed to a computer based training method. The comparison also shows if there is any value or new discoveries that can be found when using the computer based training. To analyze this data, the groups were split up among the two modalities and their abilities in reenacting the cases was based on how they were trained and evaluated (Errichetti, 2006). This group consisted of forty medical students. These medical students were required to participate in a clinical skills assessment where they had to be involved in 5 different stations that had different requirements. Some of the subjects had a computer trained SP

while others had an SP that was trained by an instructor. This interaction was recorded and archived for further analysis, debrief, and education. Findings of the data showed that there is a larger and better accuracy of the standardized patient skills that were part of the computer-based trained group. In addition, although the traditionally trained SPs were more realistically performed, the metrics displayed that they were considered less dependable than the computer trained group. This study supports that the development of an integrated training process that is created on a computer based training medium can allow medical students to work with standardized patients who are not only more dependable but prepared. Educational programs that are delivered electronically to students can provide a benefit in learning concepts in the medical realm (Errichetti, 2006). With this information, we can realize another dimension can be introduced to standardize patient training. The integration of these dimensions provides a solution that is more powerful than what currently exists. Below is a comparison of the different modalities from a literature study.

Table 1: Training Modality Comparison Matrix

	Reading a Scenario	Training Model	Human Training/ Mentorship	Computer Based Training	Integrated Training	Avatars	Video	Audio Recording	Web-based	Didactic	Virtual People Factory	Roleplay Trainer	Inter-personal Simulator	Experiential Learning	ECAs	Team Based ECA	HCDCM
Churchill, 1999																	
Conkey, 2010																	
Errichetti, 2006																	
Furman, 2008																	
Halan, 2010																	
Kobak, 2006																	
Kolb, 2001, 1984																	
Konkle, 2002																	
Kurzweil, 2003																	
Lee, 2006																	
Lester, 1997																	
May, 2008																	
Palathinkal, 2010																	
Poggi, 1991																	
Tamblyn, 2010																	
Turner, 2006																	
Rickel, 1999																	
Rossen, 2010																	

Crossover study between Standardized Patient and Web based training

Another study is conducted as a crossover study to compare results of web based training and traditional standardized patient training. Within the last 10 years, there have been larger needs for standardized patient training (Turner, 2006). The United States greatly utilizes these approaches to get medical students, physicians, and healthcare advocates the experience they need. The SP provides a very valuable tool in clinical skills for physicians (Turner, 2006). There is a cost savings aspect to this as a standardized patient encounter is a more economical choice when compared to the faculty-led pursuits.

The article written by Turner supports that the methods of teaching clinical reasoning continues to evolve. Although there are still good approaches, there are still opportunities to improve the process (Turner, 2006). When comparing Standardized Patient (SP) training to Web-based (WB) training, SP training is proven to be more effective than lectures, and Web-based training is a little less clear in its consistency. One study showed WB training to be beneficial for some physical examination skills (Turner, 2006). However, there were four other tests that showed no correlation to the modality of training (Turner, 2006). However, some things that have been considered an advantage to using the web-based training tool was being able to train several times, a shorter learning curve, and data that updates often (Turner, 2006). In addition, a learning management system could be added to provide for a better ease of use in customizability in content. The clinical portion of the step 2cs exam has caused more of a need in standardization. However, as previously stated, there have not been too many studies which compare the results from training two methodologies of training (Turner, 2006).

A study is performed at the University of Nevada School Of Medicine. It consists of fifty-four students in their second year at the college of medicine. A randomized

computer number generator is used to have the groups split into two sections. These two groups are known as “Cohorts 1 and 2”. Cohort 1 involved the subjects to first go through some WB instruction and then have SP instruction (Turner, 2006). The second cohort was called “Cohort 2”, which first had SP instruction. This Cohort was later followed by WB instruction (Turner, 2006). Both tracks of training consisted of two sequential learning modules. The first session was a case consistent of abdominal pain, and then the case to later follow was involving a headache (Turner, 2006).

The WB case had a software package that was developed called: DxR Clinician® (DxR Development Group, Carbondale, IL), which was part of the modality. Students would begin their exercise by logging into a computer where the program would begin the cases with either the abdominal pain or headache scenario (Turner, 2006). An administrator was given authority to facilitate access of training material restricted to only randomize subjects during a 45 minute period of time.

The SP learning tasks were specific to particular medical scenarios (Turner, 2006). After each learning session, a medical educator would provide five minutes of feedback which was delivered via vocal feedback (Turner, 2006). The focus was to make both modalities as comparable as possible (Turner, 2006). The SP actors also were expected to be trained in a consistent way with reliability (Turner, 2006). The next page provides a flow-chart which compares both modalities and how they are paralleled.

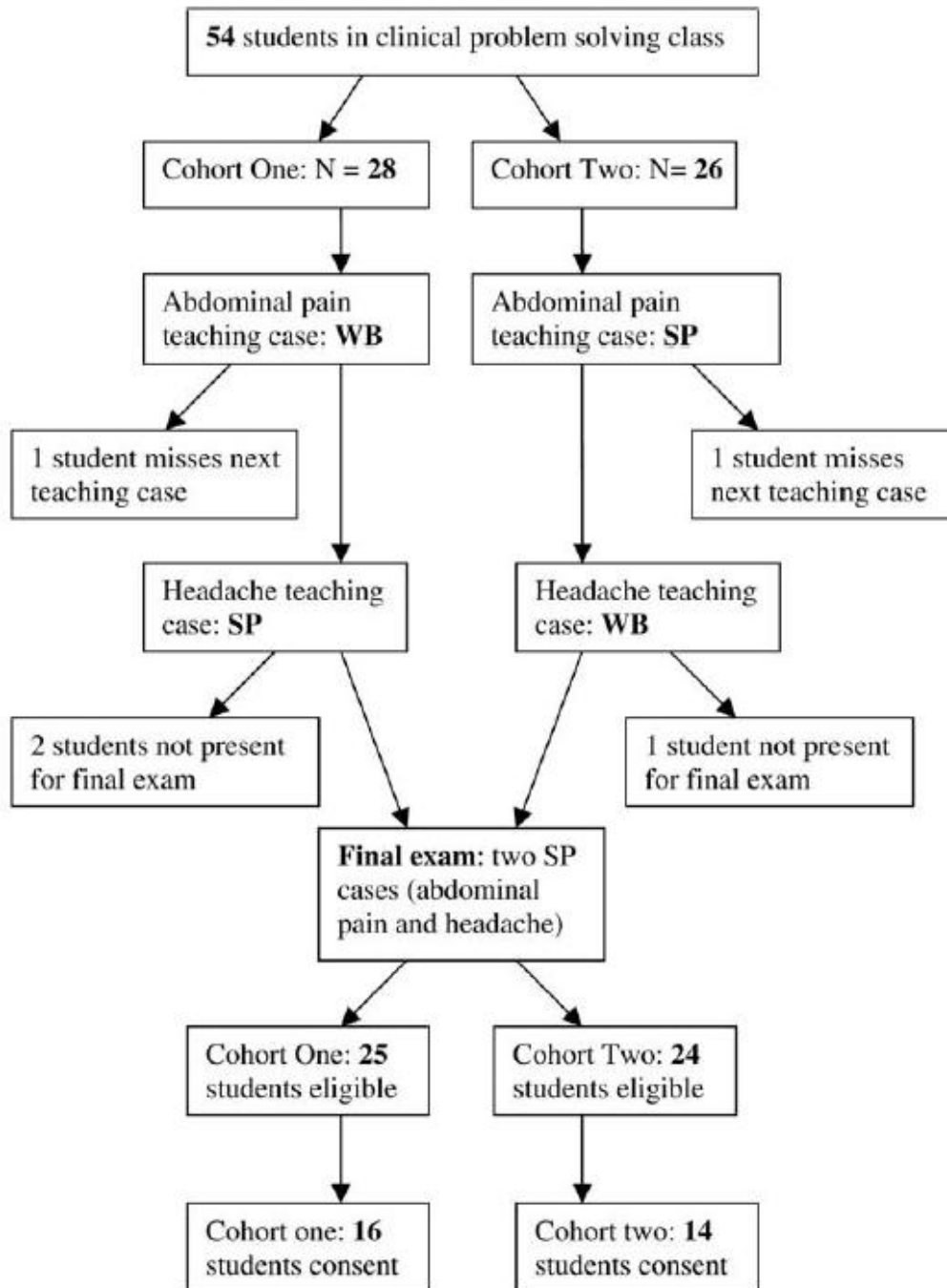


Figure 2: Study Flow Diagram

Checklists are used by standardized patients for rating medical students on their performance. For the SP cases, a physical checklist is used by faculty and physicians, while the WB modality utilized a computer. Both are good options, but this study shows which a better application is. Four days afterwards, both assessments have information to debrief the experience. This debrief includes an assessment plan, the checklists, and a

breakdown of the differences in diagnosis. After four weeks, both cohorts are required to be involved in a follow-up exam. These results show if there is a learning transfer for the concepts of clinical reasoning. The exam takes 20 minutes to complete, and during this time, faculty members are involved in observing and evaluating the performance. The graded scores are based on blinded allocation of the cohorts (Turner, 2006). A Cronbach's alpha is used to analyze SOAP (subjective, objective, assessment, and plan) notes and checklists. The alpha is utilized often to verify that the test scores are within a favorable range (Chronbach, 2009). When each teaching session is completed, the students complete a text survey on the experience (Turner, 2006). The data is captured with the use of a Likert-Scale survey. The range is from a number 1 which is annotated as a strongly disagree to a number five which resembles a strongly agree (Turner, 2006). Costs are accounted for with regards to faculty and technical support costs which were approximately fifty dollars an hour (Turner, 2006). Furthermore, there are ongoing costs depending on the length of the studies and follow-on advancements which conclude to approximately \$45 per an SP case and \$30 for the WB case formats (Turner, 2006). So the upfront costs are cheaper for the SP, but the ongoing costs are cheaper when heading to a WB platform. The statement can be made that some types of web-based training incur more additional upfront costs for development and deployment; but provides cheaper ongoing costs.

When analyzing the statistics, the tests are utilized for the purpose of analyzing SOAP note and checklist scores. Statistical significance is demonstrated if the probabilities of values are less than .05. As a result, a Cronbach's alpha of .69 is provided and a .68 for the SOAP note is given. Both cohorts have similar consenting rates, and unfortunately do not show statistical significance. The p value is .06, which is a derivative of 75% of the sample being nonconsenters, and only 47% being consenters.

What can be concluded from this is that learning with both WB and SP methods can provide grades that are within a close range of diagnosis for abdominal pain and headache SOAP notes. When considering the WB modality, there was a better grade when it came to testing of the abdominal pain but not when analyzing the headache.

Medical Student Performance with rich media

Various cases can show approaches that lead to results in performance and statistics. This case compares and analyzes the modality of implementing a rich media web module for medical training. It is not directly used for standardized patient training for the step 2cs exam. However, this rich media module is actually used for better awareness for dangerous drinking (Lee, 2006). A standardized patient is used in comparison with the traditional lecture (Lee, 2006). There was a web based training session that was used to deliver pre and post tests, Flash©, and text didactics (Lee, 2006). There were videos on two alcohol cases. Both of these videos were comparisons of a beginner and experienced examiner (Lee, 2006). A primary purpose of this training is also to enhance communication skills. There has not been an extensive amount if any of published studies on this topic (Lee, 2006).

Along with these communication skills, students are taught various topics that are in relation to the subjects of public and behavioral areas. The different methods in training are inclusive of lectures, web modules, and course syllabi. These methods assist in preparing the students for faculty-facilitated seminars (Lee, 2006). The three current parts to the curriculum consist of a lecture, small group seminar, and an OSCE (observed structured and clinical evaluations case).

For the study, the first class of 2006-2007 medical students was chosen as a cohort. The sample of N=163 was assigned to participate in a web module which was paired with a one hour lecture. However, an interesting point to observe is that both the

web-based training and the lecture were optional (Lee, 2006). Within 3-5 weeks, the two groups joined meetings with a patient that is involved in hazardous drinking; which later had an OSCE alcohol case (Lee, 1006). Demographic data was not collected as a result to restrictions of the NYUSOM IRB as the population did not have a significant amount of diversity.

The web based module that was preferred, was java-based. This was developed by the NYU Division of Educational Informatics faculty. The website can be accessed at the url: <http://edinfo.med.nyu.edu/alcoholscreening/> (Lee, 2007). The login is the word: demo; and the password is: demo) (Lee, 2007). For the duration of the week, students were allowed to use an ID and password to login to the module where they would agree to submit their responses for the contribution of research. The pretest consisted of multiple choice questions that ranked answers based on a scale (Lee, 2007). This was used to get the learner accustomed to educational information along with understanding gaps in knowledge (Lee, 2007). The knowledge sections included safe drinking limits and also provided background information for two alcohol vignettes (Lee, 2007). In addition, there was a separate area which had an outline that decomposed the learning modules. A voice over provided the feature to have audio narration. This also keyed in on particular areas of interest. Many of these highlighted principles used the applications of Flash© animation and hyperlinks to external sources (Lee, 2007). This was a great cost savings application as it reduced the need for extraneous text which could be leveraged from other sources. Two video clips were used to illustrate teaching and training as part of the curriculum (Lee, 2007). One of the videos consisted was a male with pancreatitis who is attempting to stop drinking and the other video has a person of the opposite sex who fails to follow a reasonable amount of alcohol levels and later decreases the amount that is consumed (Lee, 2007). Later, an inexperienced student's

interview was followed by a physician who was in the field for a long time which had expert examples of interview techniques (Lee, 2007). After each 1-3 minute clip, the learner responded by typing in an area to input the text. Afterwards, the responses to the learner responses were coupled with the responses of the teachers (Lee, 2007). When this was complete, there was a posttest which had the same pre-test items; comparisons of the answers, and finally a module feedback prompt (Lee, 2007). The time duration of this was approximately 45 minutes. The lecture had the same steps for the procedures (Lee, 2007). Afterwards, there were small group seminars (Lee, 2007). Three to five weeks after the web and lecture training, an OSCE using SPs with two cases was put in place (Lee, 2007). Female SPs were educated and trained on roleplaying a woman character that reduced the intake of alcohol and was later evaluated and assessed individually (Lee, 2007). The SPs did not know if they were either lectured or WB (Lee, 2007).

The study had a sample of 82 subjects which was compared to the sample of N=81 groups who studied a lecture. A similar amount of students signed up for the web-based in comparison to the lecture training. It was 82% to 72%.

There were reasonable results from this trial version of a web based training module. This web-based training represented a great tool for intervention of hazardous drinking. It was a great training for clinical skills for medical students. A significant amount of the students chose to use the web-based training as opposed to the lecture version of instruction. 5% of the lectured students were defected to the web training as a result of course instructions (Lee, 1008). 6 out of the 75 had some technical difficulty when using the computer. Constructive criticism from the students allowed the maintenance of a list of processes that could be improved (Lee, 2008). This was especially significant when expressing that the students preferred the web based training in comparison to the lecture (Lee, 2008). The web based training results were just as

good as the lecture based training (Lee, 2008). There were better assessments of a patient's hazardous drinking.

However, there are some things that can be improved with the system (Lee, 2008). Presenter bias may have happened as a result of web authors being involved in the training sessions (Lee, 2008). SPs were expected to have higher reliability, but they occasionally overestimated student performance (Lee, 2008). Finally, some of these results were not beneficial to the students but show that web based training can be analyzed to ensure applied measures for clinical skills (Lee, 2008). These applications should be justified for cost and effectiveness (Lee, 2009). In conclusion, this is another instance where web-based training provides a positive result for training and automated assessment. Web based training is chosen as a more preferred medium when compared to traditional lecture based training. Not only is it a preferred choice, but it is more convenient for students.

Rater Training

Rater training is important when trying to understand how to rate medical students and assess their performance. Rating is crucial when understanding a standardized patient's responsibility to properly gauge a medical student's competence. Peggy Wallace, the author of "Coaching Standardized Patients", recommends to always request the SP to fill out each checklist immediately after each encounter. The most accurate time is to fill out the survey when the scenario is fresh in the SP's mind (Wallace, 2007). This is an effective approach as there are several occasions where there may be a specific action that the SP performed that was not connected to every checklist response. An example of this is when an SP forgets if a student palpates the lower right quadrant of the abdomen or not. If this occurs, she might recall if she gestured for the pain during the actual assessment. That is why it is crucial to have each standardized patient's response

integrated into the history (Wallace, 2007). If this is done, then the patients are more inclined to respond regarding to how they felt the student as opposed to whether actual items on the checklist were completed as a result of not recalling the scenario (Wallace, 2007). Not being able to recall things properly or having a bias can also cause inaccurate ratings which can cause inaccurate information (Wallace, 2007). Inversely, if there are actions that affect reasons for negative biases, data can be misrepresented as well (Wallace, 185). Many errors in SP rating also are contributed to faculty recall. Some of the events in multiple encounters begin to mix with each other. When filling out rating assessments there are occasional instances where this is confusion if a particular behavior was from a current scenario or a previous one (Wallace, 185). Some other strategies are to make sure the SPs complete all checklist items before writing in all their comments and letting SPs know which checklists are likely to be most difficult to accurately complete (Wallace, 185). The following pages elaborate on sources for unreliability and bias in rating and different modalities and effectiveness in rater training methods.

Rater Rapport

Understanding the user-centered and relationship based approach to clinical skills and healthcare is not always clearly defined. As a result of the ambiguity in this area of communication skills, relationship building, and therapeutic alliance; there has not been much research that backs that the fundamentals of this concept are effective (Hall, 2009). This study shows what happens when the relationship of students and SPs are evaluated by subject matter experts. Analogue patients were also used who were individuals who viewed the videotapes and imagined themselves being the patient as another rating tool.

One of the first visions that are accomplished in rating medical students is the degree of rapport that is seen (Hall, 2009). A second step is to have an understanding of the things to come that are related to rater rapport (Hall, 2009). Having this foresight is

crucial to preparing for an interaction that is highly and lightly involved. A third focus area is to be able to determine and realize the differences of validity when it comes to a single 1 minute excerpt that is compared to several 1 minute snippets of media (Hall, 2009). Rapport is considered a wanted feeling that is shared by both parties that are involved in the interaction (Hall, 2009).

Interrater Reliability

Inter-rater Reliability is important when using different raters for evaluation of a standardized patient encounter. There are several different aspects that need to be analyzed. Inconsistencies can be possibly seen in areas where raters are required. Biases and standardization are issues that won't always be consistent, but results can be found out when studies are performed. The reliability studies in the following sections show examples of how inter-rater reliability is crucial in properly rating medical students for high stakes exams and clinical skills assessments.

Bias and Poor Interrater Reliability

An examination that focuses on neurological clinical skills is analyzed for bias and extent of inter-rater reliability. Residents were graded based on certain qualities and skillsets with the use of a Clinical Evaluation Exercise (Mini Cex). Some of the things in particular that were looked at were clinical skills, attitudes, and behaviors (Schuh, 2008). This experiment examines inter-rater reliability along with bias that is present in a Neurology Evaluation Exercise (NEX). The local faculty that does not have the sufficient training is observed to see if it behaves different than ABPN examiners that are not affiliated (Schuh, 2008). Furthermore, some other goals in this exercise was to see if the NEX scores were within a favorable comparison with the neurology residency in-service

training examination (RITE) scores and to also see how many studies are needed for reliable testing and passing (Schuh, 2008).

Approximately four different departments were involved in the study and some demographical information was captured for research purposes (Schuh, 2008). Since it was important for the identities and experiences to be unknown, the residents were taught to be at a low competency (Schuh, 2008).

The sample consisted of thirty-two residents (21 men, 11 women) who were involved in the research (Schuh, 2008). 63% of the population, which is a larger majority, was PGY-2 residents (Schuh, 2008). The performance of the Pre-study for the RITE in 2006 was fluctuating throughout these groups (Schuh, 2008).

As the goal of the study was not to prove that the graders knew who they were grading, it was ok if some of the graders had some preferential treatment (Schuh, 2008). The more important scope of the study was that the faculty did not know how advanced or not advanced each of the residents was (Schuh, 2008). Even though the inter-rater reliability level was minimal, there was still a similar bandwidth of the range that fell within a 75 percentile performance. A videotaped session that was later debriefed to raters could be a valuable learning tool that would pose some training value. Modeling and simulation could also be used in this dissertation with applications of virtual humans and conversational modeling to replicate scenarios.

Causes for Bias and Lack in Reliability

Having set preferences can greatly affect the reliability of the rating (Tamblyn, 1991). This passage describes a study that discovers the origins of the unreliability. Two universities were involved in facilitating an SP test that is exactly the same; which was created for clinical clerks in 1987(Tamblyn, 1991). As a result, the rating bias was due to

the test site: Videos of the population sample were taken and it was seen that there was a reasonable range of reliability (.37 to .52) (Tamblyn, 1991).

The test site did not play a part in the rater agreement, but there were differences in the score (Tamblyn, 1991). Sometimes, an amount of partiality can be reduced by putting a particular framework in place to not allow the mind to give special preference to a particular treatment in a scenario. The following study delves into the comparisons of web based versus traditional rater training.

Enriched rater training using Internet based technologies

One of the important things that are required when it comes to training is producing high quality individuals who are able to train consistently and effectively. This can be done with the proper instruction. Although this need has increased, the sufficient amount of research has not been performed to show enough data to support this. Another reason is because it has not fully been implemented in the healthcare system (Kobak, 2006). When trying to gauge the performance of a rater, there are two traits that must be kept under close consideration. The first skillset that must be acquired is conceptual knowledge. The second is applied skills (Kobak, 2006). Conceptual knowledge plays a large part in understanding the way that scoring is done, how things are done to scale, and the facilitation of scaling (Kobak, 2006). However, applied skills allow the ability to display how well an individual can conduct a clinical interview (Kobak, 2006). Even though both of these skillsets are expected, many individuals simply use training literature or a pre-brief with a healthcare advocate. According to Kobak, approximately 75% of the studies use this method, and are not recommended as the most optimized solution. Proper rater training is essential to the Standardized Patient community so that every medical student has a fair advantage to being standardized in the process of the clinical skills portions of their board exams.

To provide a solution to this gap in SP education, training, and performance, a training tool has been developed. This tool allows an interactive experience, and has a didactic flow to it (Kobak, 2006). The internet and didactic instruction was used to deliver the particular modules which address the particular learning task that was needed (Kobak, 2006).

The main idea is so understand and isolate a traditional training approach with a new revamped version that not only grabs the audience's attention more, but also engages the new user in this enhancement (Kobak, 2006). One of the objectives is to see if there is a benefit before and after the training when applied conceptually (Kobak, 2006). The expected result is that when training is enhanced, it is more effective than traditional (Kobak, 2006).

Subjects were recruited to participate in a study (Kobak, 2006). There were a total of 22 sites where everyone participated in (Kobak, 2006). Out of these sites, there was one that declined, and 2 of them that did not have the internet connection that was feasible enough to handle the technical requirements that were specified (Kobak, 2006). 12 of these sites were blindly chosen to be involved in the experiment out of the remaining 22 sites (Kobak, 2006). Out of the 12 sites, six of them were used to be part of the traditional training and the other six were used for the enriched training (Kobak, 2006). Both of these training modalities were completed within one month, so that there was no negative training for the duration it took to complete (Kobak, 2006)

The learning task that was created by the sponsor for the traditional method was to be able to train raters for the HAMD (Kobak, 2006). These tasks consisted of the following: an independently scored video tape, statistics were analyzed from the scores, a general overview, a video tape was shown and the sponsor reviewed and assessed the

clinical interview proficiency of raters and gave inputs to the research group (Kobak, 2006).

For applications of this study, the word “enriched” translates as web-based training with remote, didactic, and applied capabilities (Kobak, 2006). Within this training, capabilities to send emails to the instructors along with print for the text of the teaching modules were also available (Kobak, 2006). The didactic section took about two hours to complete, which was later followed by a clinical training section. During this training section, one interview was a patient, while the other was a standardized patient. Since these interactions occurred via teleconference, there was a larger area for evaluation and measurement. The final portion allowed different students to exchange words with different patients and then get together for a close-out discussion session.

The results showed that the traditional method of training did not show significant effects in the areas of conceptual knowledge or clinical skills. Along with the interaction, part of the reason for there being an effect in enriched rater training might be a function of time (Kobak, 2006). To elaborate, the pace of training that was given within the given due timeframe reduced the amount of pressure to be complete in a certain period of time. This provided additional time to be able to train, get up to speed, and have sufficient practice (Kobak, 2006). Web-based training could also be considered beneficial as it would allow trainees to login into the system when the time was convenient. Virtual classroom interfaces could be used to provide a method of training that is more distributed via the internet (Mott, 2000). However, the ability to have teleconferencing features allow providing flexibility in locations. When comparing this to live training, there are more opportunities for standardization and collaboration for training pursuits (Kobak, 2006).

Avatar Immersion and Presence

Avatars can be integrated into innovative and collaborative training solutions that can allow for an experience that brings them into a specific environment that is related to the training task at hand. They have the ability to bring the students into an environment where a new awareness of occurrences could happen (Privateer, 1999). During extensive periods of time, studies have applied concepts that are augmenting teaching with simulations that can result in increased learning and retention (Woodward, Carnine, & Gersten, 1988). One specific finding for students with low reasoning ability was that simulations offered a bridge to learning and understanding (Cox, Abbott, Webb, Blakeley, Beauchamp, & Rhodes, 2004). Simulations also show a creation of interaction and reflection opportunities for learners working with models (Kolb, 1984; Kolb, Boyatzis, & Mainemelis, 2001; Montgomery, Brown, & Deery, 1997).

Immersion is when many of the senses are taken over with simulation, and the user feels as though they are part of the experience and the aspects of the simulation are a normal extension of their functions even though they are in a simulated or in a synthetic environment. In addition, when a user is able to interact with a mediated interface, it could provide a sense of immersion.

Presence is also known as having the sensation of being at a particular location or experience (Heeter, 1992; Lessiter et al., 2001; Witmer & Singer, 1998). During this experience a user is able to participate in an experience that has a realistic look and feel. Furthermore, they react when there is an activity that captures their attention (Heeter, 1992) Virtual Reality has been involved in stimulating the stimuli of the human mind to bring in a certain level of presence.

Facial Expressions in Communication

Humans are able to communicate with their face and language (Cassell, 2000). However, when combining facial expressions and gestures, new meanings can be manifested. Some aspects that are interesting when it comes to face-to face conversation are facial expressions, gestures, and immersive experiences (Nickerson, 1976). These communication aspects are things that can also be integrated into a human computer interface. An embodied conversational agent (ECA) would incorporate body movements, emotion, and model the functions of conversation that underlie the features. It is also important to understand the traits and reactions of a user to be able to properly model them. One thing that allows an avatar to relate to a homo-sapien is being able to recognize the normal cues that an individual produces (Cassell, 2000). One example is when an ECA can serve as a tutor that can teach new lesson to a user, but still show relation and empathy towards the user's motives.

A performative is a key indicator of intent of a speaker (Cassell, 2000). Furthermore, the performative suggests why the speaker is communicating when it comes to a specific topic (Cassell, 2000).

Facial displays can be used to indicate a certain feeling, and when expressed a certain way can also take the place of certain words. Eye gaze is also an important tool in facial displays as they are able to control the level of conversation. Computer Science groups have been able to create synthetic representations of these facial expressions and gestures in ways to assimilate to human behavior (Cassell, 2000). Gestures can be unexpected and are categorized as: Iconic, Metaphoric, Diectic, and Beat (Cassell, 2000). Iconic gesture can resemble a particular perspective (Cassell, 2000). Metaphorical gestures can resemble a particular metaphorical symbol (Cassell, 2000). This type of symbol allows a better understanding of a concept such as a person rolling their hands

together while saying the phrase “get the ball rolling” (Cassell, 2000). Diectic gestures take up a particular area and solicit cues. Beat gestures are small “baton like” motions can get the attention of a user and also provide opportunities to symbolize particular aspects of the conversation (Bavelas et al. 1929).

The timing of different movements in the eyebrows and forehead can also play a particular role in accenting the importance of certain words and messages. Some of these gestures have multiple functions (Cassell, 2000). For example, a smile could indicate happiness, but could also provide a good medium for social interaction.

Designing Embodied Conversational Agents

One of the most important differences between an ECA and an avatar is that an ECA has a conversational model and is created to be more conversational in its nature. An ECA also has the ability to recognize particular cues (Churchill, 1999). A good natural conversation that an ECA has, should be built on the behavior of a human-like interface; and a good indicator of its intelligence is its ability to engage a human (Churchill, 1999). An example of an interesting observation is if an ECA were to ask a question where the eyebrow would raise and the human would respond with a headnod as opposed to speaking a response (Churchill, 1999). Even though, two conversations are never exactly identical, there are common habits in a conversation that be used as a good path for conversation like: interaction, responding to queues, waiting for a cue to speak, inflection, emphasis, and breaking away (Churchill, 1999). A framework that is static but also dynamic allows the structure of the conversation to be flexible in its ability to interact with a user (Churchill, 1999). The system should also ensure that the ECA gives an opportunity to take turns in its interactions and also focus on the conversational functions as opposed to the mere sentence (Churchill, 1999).

An example of an ECA is provided below where a conversational model was developed to give a user a tour of the house (Azerbeyajani, 1996). Inputs from different devices were procured with the use of input sensors (Azerbeyajani, 1996). The software that was used was called STIVE vision which implements the use of recording devices to capture the realistic aspects of the images (Azerbeyajani, 1996). Data that is received is given a time stamp for documentation purposes and then processed into a module that analyzes the understanding of the information (Churchill, 1999). It later collaborates with an understanding module to determine the state of the conversation (Churchill, 1999). In the example below, the model's name is REA. This model knows to generate a response after a speech act and gesture is made (Churchill, 1999). The generator that is used to understand the objectives of communication from a request formulator is called a SPUD generator. It uses basic knowledge that is programmed to determine the premises for interaction (Churchill, 1999).



Figure 2.4
Rea says, "It is next to the bedroom"

Figure 3: Rea Model

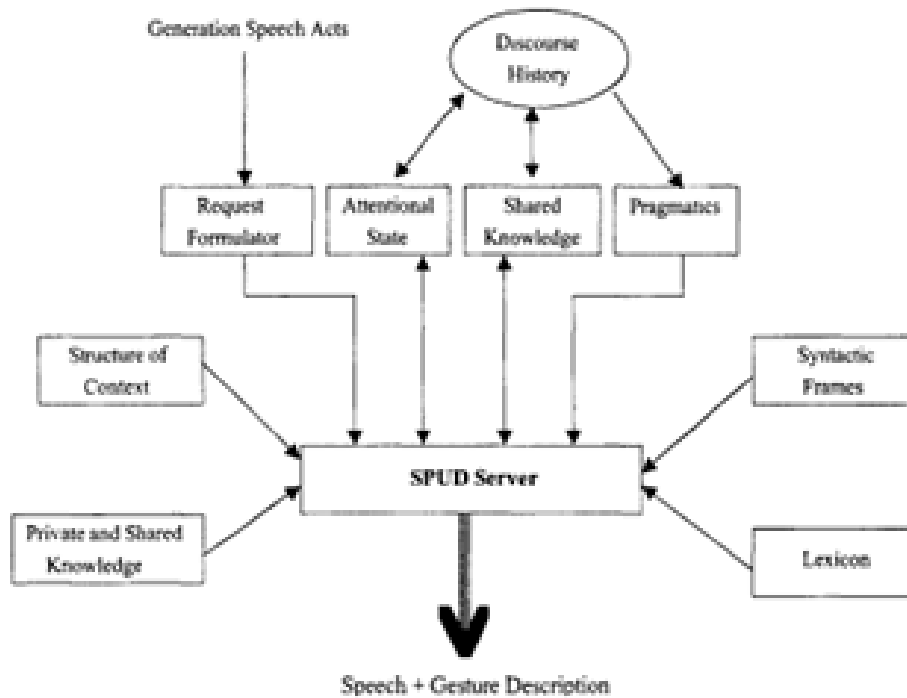


Figure 4: Speech and Gesture generation with SPUD server

Along with an input system, it is also essential to have an effective output system so that a good flow of conversation is framed in an organized model (Churchill, 1999). Below are some diagrams that display these layers of conversational outputs in a model; along with a diagram that shows how synchronized speech is used in an action schedule (Churchill, 1999).

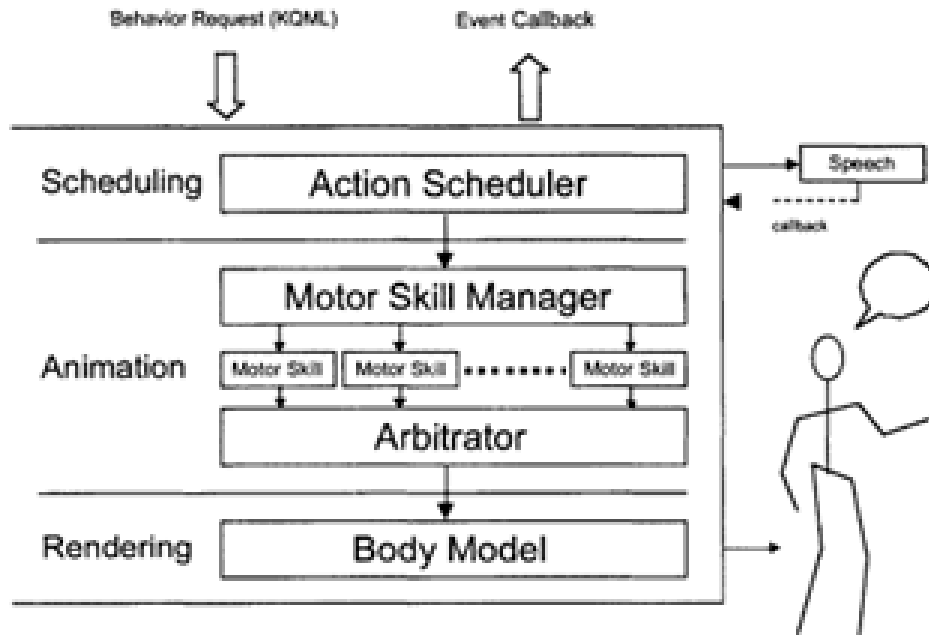


Figure 5: Three layers: scheduling, animation, and rendering

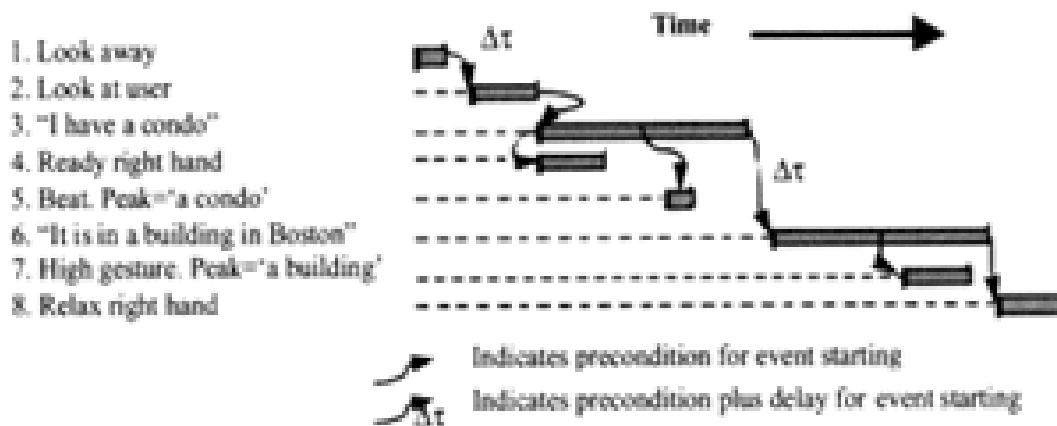


Figure 6: Synchronized speech with the action scheduler

Embodied Conversational Agent Allies

Different aspects of conversational modeling should come together at the end to accomplish a common goal and vision. Scenarios and storyboards can be used to properly plan how ECA interactions can effectively occur (Cassell, 2000). The process can evolve as the general design could be vague, but then more low level details can be elaborated on (Cassell, 2000). These storyboards and general designs can later be

converted into an initial prototype of the conversational model (Cassell, 2000). Some things that are also laid out of the agent in the design are the ECA's personality and nature of interaction in a conversation (Churchill, 1999). The scenarios also expect tasks from the ECA; and the EDA should also be pleasant to look at (Churchill, 1999). The ECA's can sometimes resemble a human, but does not necessarily always have to (Churchill, 1999). Abstract shapes can also effectively take the shape of an ECA (Churchill, 1999). Below are some drawings that investigate the issues in design for ECAs and display the architecture for ECAs.

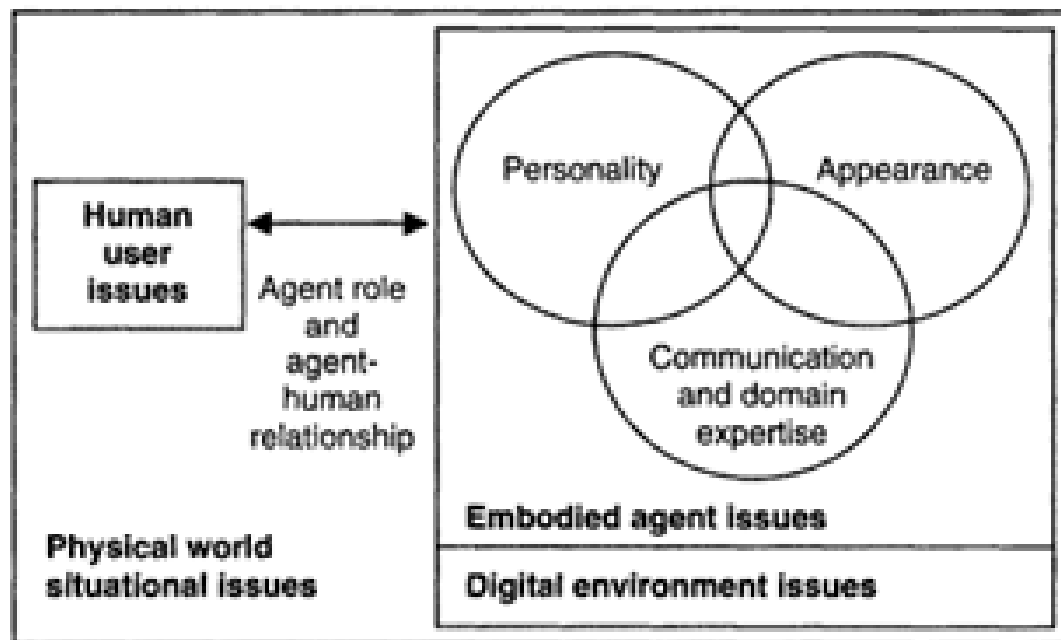


Figure 7: Issues in design for ECAs

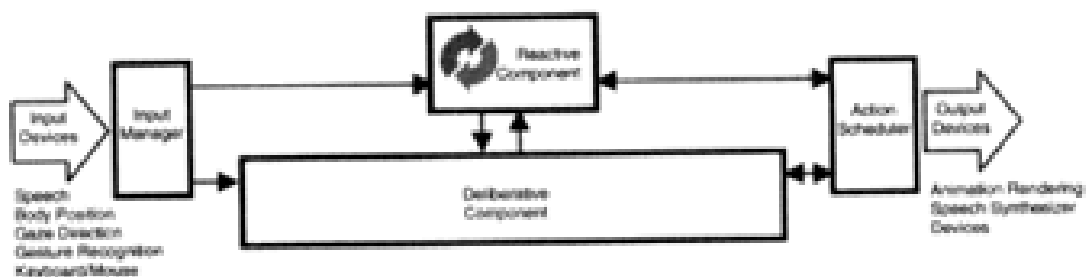


Figure 8: Top Level ECA Architecture

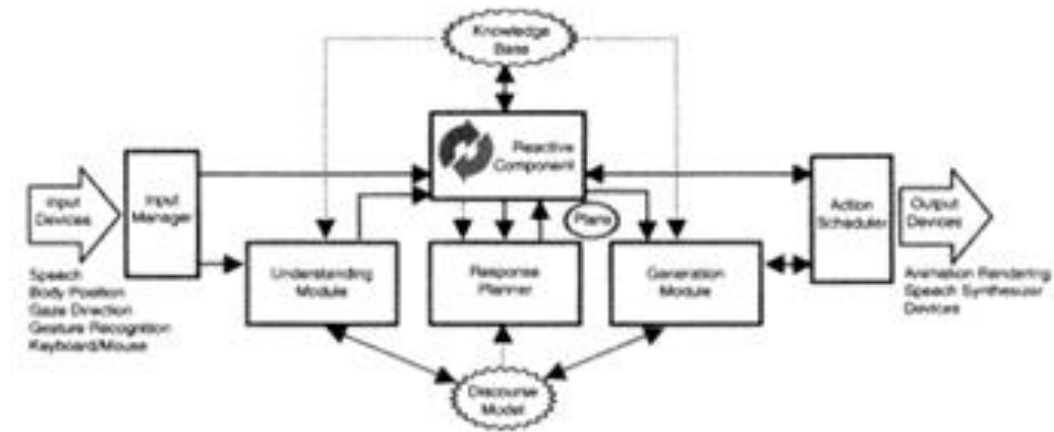


Figure 9: FXPAL ECA architecture

Team Based Training

Team based training is critical when it comes to training objectives. Using a distributed system in a virtual environment like in Second life can create this type of scenario. Along with being an instructor in a virtual world, an instructor can also call on a colleague that can also assist in co-instruction for a particular class. Below is a training session with an avatar named Steve who provides training on how to extinguish an alarm (Rickel, 1999).

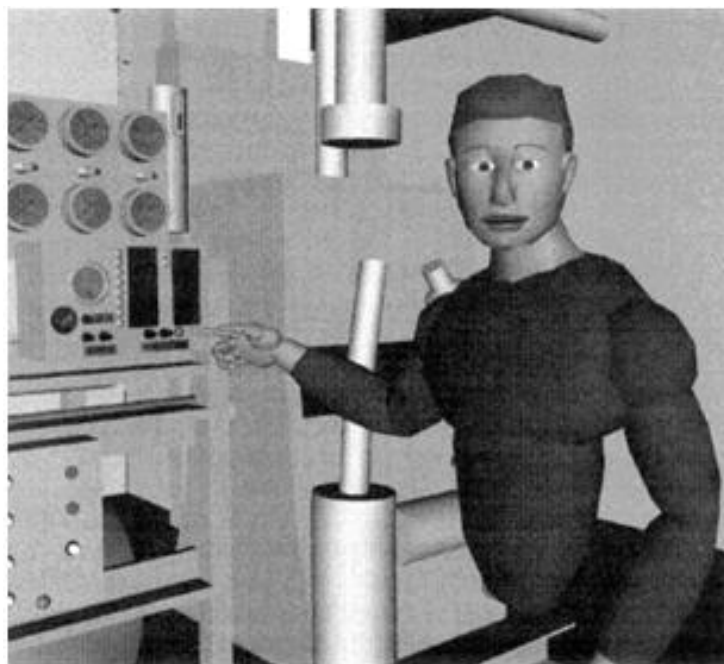


Figure 10: Training to Extinguish an alarm

This virtual scenario was created by Lockheed Martin to show how an ECA can assist in teaching to learn a particular task (Johnson, et, al, 1998). There are a few things that are essential for creating an immersive world. One of the initial things that are needed is an effective simulation. Simulation is important as it controls the behavior and feedback of what occurs in the virtual world such as VIVIDS which is discussed by Munro et al, in 1997. In addition, a visual interface that runs over the engine but be present and also pleasant to view (Rickel, 1999). The requirements that are established up front to determine how realistic the simulation is, becomes one of the determining factors when investigating how effective a virtual world is. Or a simple computer screen can be used of the requirements are simplistic. The recognition of speech is accomplished, but a system that comprehends the signals of speech (Rickel, 1999).

Cognition in conversational modeling consists of three layers. One of the lowest layers would model general human intelligence (Newell, 1990). Tasks would be defined more in the second layer, and the third layer would categorize the knowledge within the domain. A domain can be updated to accommodate different learning tasks, needs, and lessons (Rickel, 1999). The diagram of perception, cognition, and motor control can be seen in the diagram below.

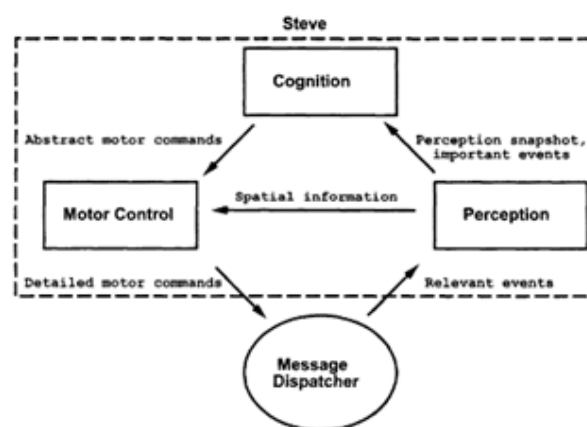


Figure 11: Cognition, Perception, and Motor Control

The task oriented module is where the collaboration occurs for different tasks and objectives (Rickel, 1999). In order for some of the events to occur in a natural fashion, the ECA may not always have complete control of the actions (Rickel, 1999). Semi-automated systems could assist in managing the overlap where an autonomous entity is needed.

Behavior is controlled by rules: Operator proposal rules are used to assist in setting the configuration for the users (Rickel, 1999). When implementing agents that are lifelike, good pedagogy is displayed (Lester et al 1997b). However, there are two issues when considering quality in instructional value. The first is diectic believability which is having the agent realize the physical properties of the environment that it exists in. The second is emotive believability which refers to the agent being expressive and also motivational to the user (Lester et al 1997b). Diectic frameworks are used to address these two aspects of believability (Lester et al 1997b). A prototype was created to test the effectiveness of a Diectic framework called Cosmo (Lester et al 1997b). This agent teaches network routing mechanisms with the use of Deictic gesture, locomotion, and speech (Lester et al 1997b).



Figure 12: Deictic test bed with Cosmo the interactive agent

The educational benefits to emotive behaviors are as follows: the agent showing empathy in the educational pursuit and a sense of companionship, cognizance of the student's performance and enthusiasm for the participation (Lester et al 1997b). The student's problem solving activities drive the engine for this (Lester et al 1997b). Some of these behaviors can be constrained within particular domains, can be triggered based on particular actions, and can be customizable as the pursuit continues. Below are diagrams of architecture for Agent behavior planning and anatomical emotive information.

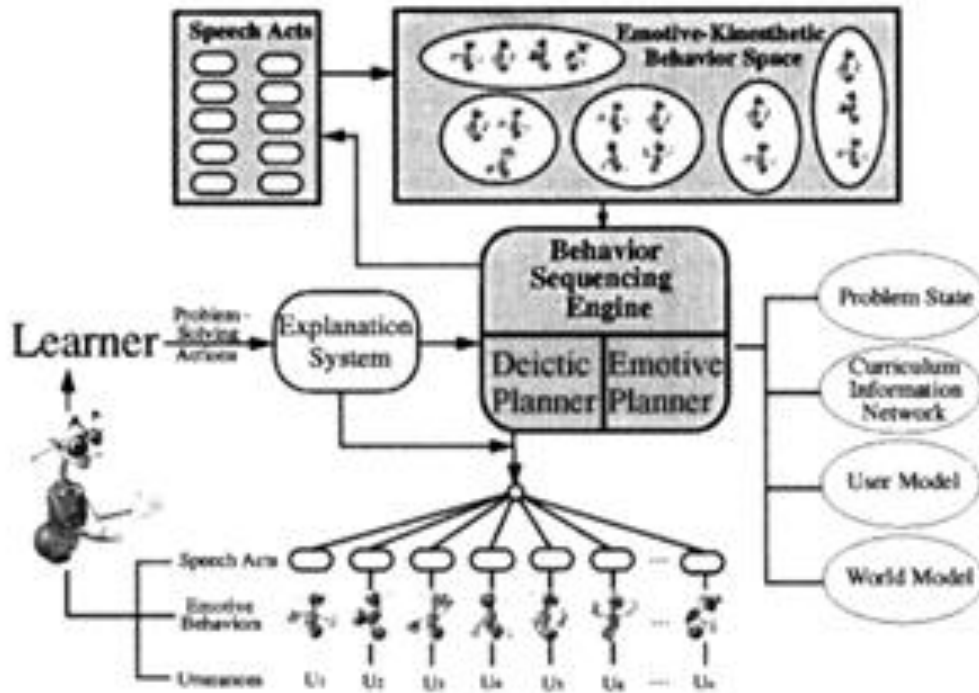


Figure 13: Agent behavior planning architecture

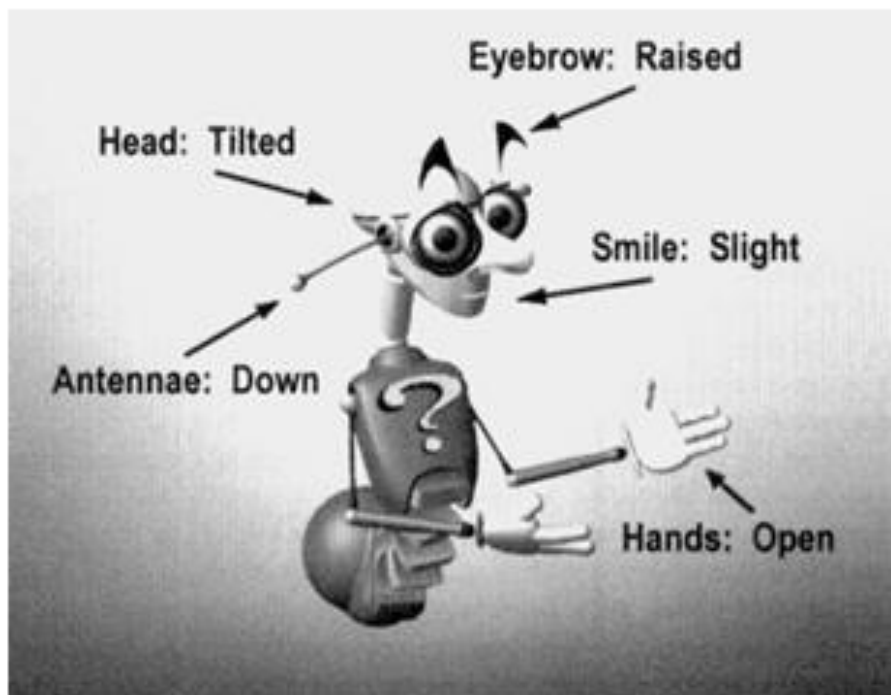


Figure 14: Anatomical Emotive Information

Communicative Tasks

Communicative tasks can also be bound by the level of formality that is expected in an interaction (Poggi, 1991). These factors can also be driven by the expected goals and emotions from the conversations (Poggi, 1991). Below are some examples of expressions that are used to compliment communicative tasks.

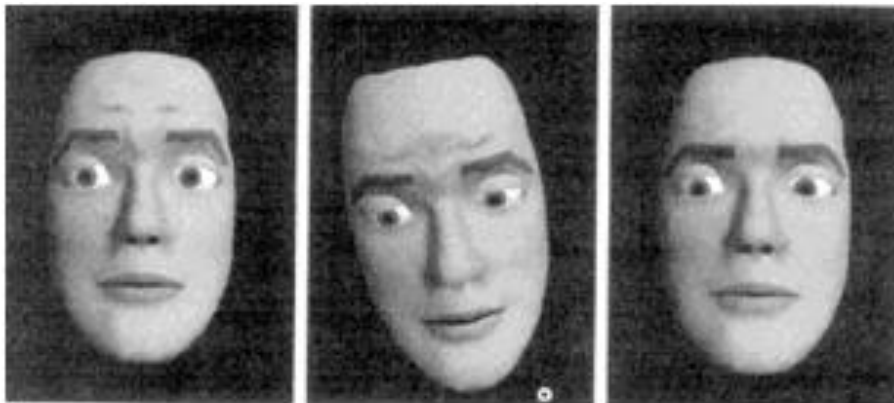


Figure 15: Expressions- Suggesting, imploring, and ordering

Embodied Conversational Agents and Simulated Personalities

Computers eventually serve as companions and assistants, and the extension of a mouse, keyboard, or handset could be obsolete (Ball, 1997). “An assistant” would be able to effectively comprehend different ways to deliver the same message and also be able to provide feedback (Ball, 1997). There are several instances that a human expects the same type of feedback from an ECA that it would from a human (Ball, 1997). Understanding the social impacts in conversational modeling assists in guiding the expectations that humans should have when interfacing with ECAs. However, conversational modeling in the ECAs can also improve to make this adjustment easier in transition. Emotion can be modeled in the ECAs so that they may accommodate different stimuli from the users (Ball, 1997). However, emotional and expressive behavior is still being developed to be successfully simulated in a high fidelity arena (Ball, 1997).

Emotional responses are established in two dimensions (Lang, 1995). Valence defines the positive and negative aspects of feeling, while Arousal displays the intensity that is involved in a response. The figure below shows how these two values are represented and can be cross referenced with each other depending on what particular emotion is of interest. The second figure displays the representation of friendliness and dominance.

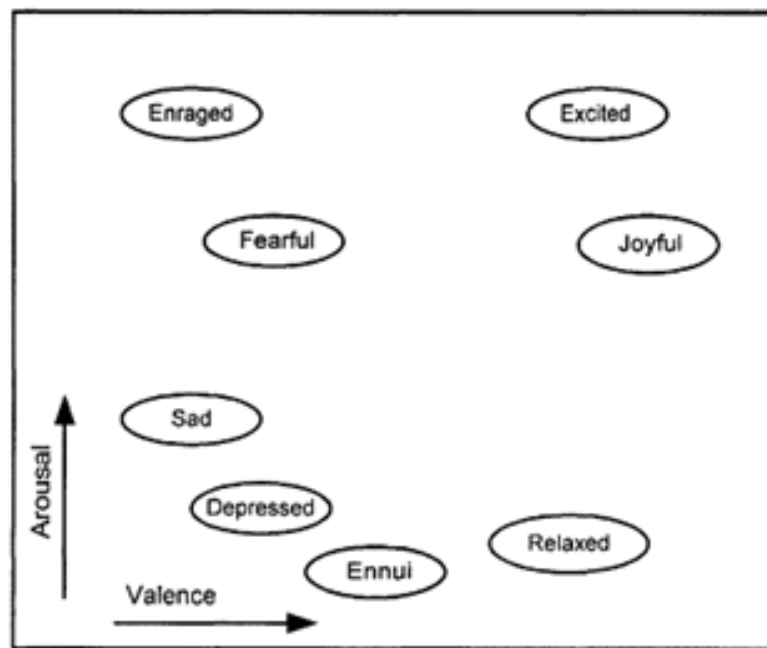


Figure 16: Position of emotions within Valence-Arousal space

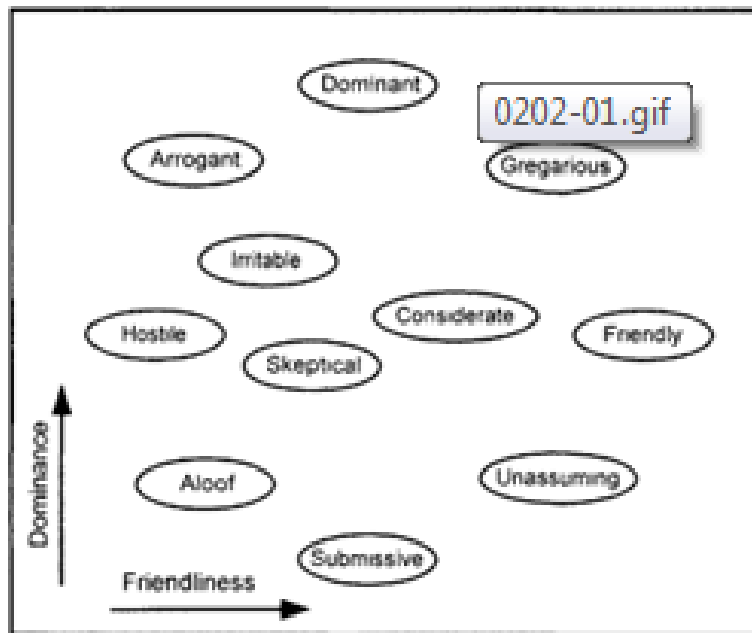


Figure 17: Position of personality within Dominance-Friendliness space

A Bayesian network can be used to model emotion and personality (Ball, 1997). This type of network is able to predict the potential for different outcomes (Ball, 1997). It is also able to display an organic representation of computer behaviors that are within a particular emotional state (Ball, 1997).

A psychologically aware system is able to analyze characteristics of the user and is also able to make judgments (Ball, 1997). A model of this is able to do the following: observation (study the sensory input), assessment (using an algorithm to determine performance), using a policy component to determine relations, and behavior of the simulation (Ball, 1997). The policy module makes judgments of the user's psychological state by analyzing the conversation (Ball, 1997). Emotionally diverse phrases must also be accommodated using this concept (Ball, 1997). Vocal expression carries a stronger influence than the text based expressions that are delivered (Ball, 1997). Inside the network is an expression node which provides a relation to arousal, valence, and

friendliness (Ball, 1997). Faster body movements can also cause more emotional arousal (Ball, 1997).

ECAs that Present and Sell

Presentation software has provided the need for more integration with ECAs that are intelligent (Andre, 1995). When integrating these assets into presentation software, the look, feel, and interaction can provide for a more attractive solution to instruction and content delivery. The advantage is that there is an on demand speaker who is always available to facilitate a crowd of individuals on learning a new concept, idea, or school of thought. An example of this was created. This embodied conversational presenter was called a Personalized Plan- based Presenter (Andre, 1995). The agent did have its own personality that was autonomous such as gestures, idle time, and immediate reactions (Andre, 1995). In order to test this, a study was performed. Half of the subjects participated in watching a presentation that was facilitated by an interactive agent, while the remaining subjects watched the same module with only audio along with arrows that highlighted important areas of interest, which lacked the gestures and expression that the ECA had (Andre, 1995). However, after analyzing the data that was collected, reading comprehension was not effected; but the expression and feedback of the users had a difference (Andre, 1995). However, instilling confidence was not sufficiently instilled with solely the use of an agent. Teams of agents can also be used to teach a particular learning task or lesson. When initiating a design for a presentation, some things to keep in mind are dialog types, character roles, dialog contributions, and scripted behaviors (Andre, 1995). Agents can also enter the marketplace serving the role of as a sales person. Microsoft Agent can be used for this (Andre, 1995). Along with a salesperson, other roles can be taken on such as a financier (Andre, 1995). When developing this type of solution there are two types of personality factors that must be

created (Andres, 1995). The first is Extroversion and the second is Agreeableness.

Below is an image of agents that can be used in sales.



Figure 18: Agents used in sales



Figure 19: Assuming roles for agents in sales

Many of these dialogues can be complex at times. To mitigate this, the goals must be broken down into simpler pieces. It is the central planning component that is in charge of this. The ECA's are autonomous for the most part, but there are events that can trigger the personality traits of it being either an introvert or an extrovert (Andre, 1995).

Smart Avatars

The best way for smart avatars to be modeled to be similar to a human is to be connected to the synchronization of motion capturing technology, so that posture changes, reaching, pointing, grasping, etc...is applied (Badler, 1999). For gestures that are not similar to real life movement, alternate designs must be created.

When interfacing with an avatar, natural language processing can be used. Some words that are exchanged can be conditional based on the feedback. One example of this is if an ECA is cognizant that another ECA shows traits of being unpleasant or hostile, and knows to draw a weapon (Badler, 1999).

An example of this behavior can be seen in Jack's Moose Lodge where there are five virtual humans that interface with each other (Badler, 1999). Some of the agents are able to be autonomous, while others have been programmed with certain behaviors (Badler, 1999). Also below are some other examples of occurrences when an ECA acts upon a response. This virtual environment serves as a platform for multiple users to collaborate on a certain task, communicate, and also sense one another's intentions and act upon them with a response (Badler, 1999).



Figure 20: Jack's Moose Lodge

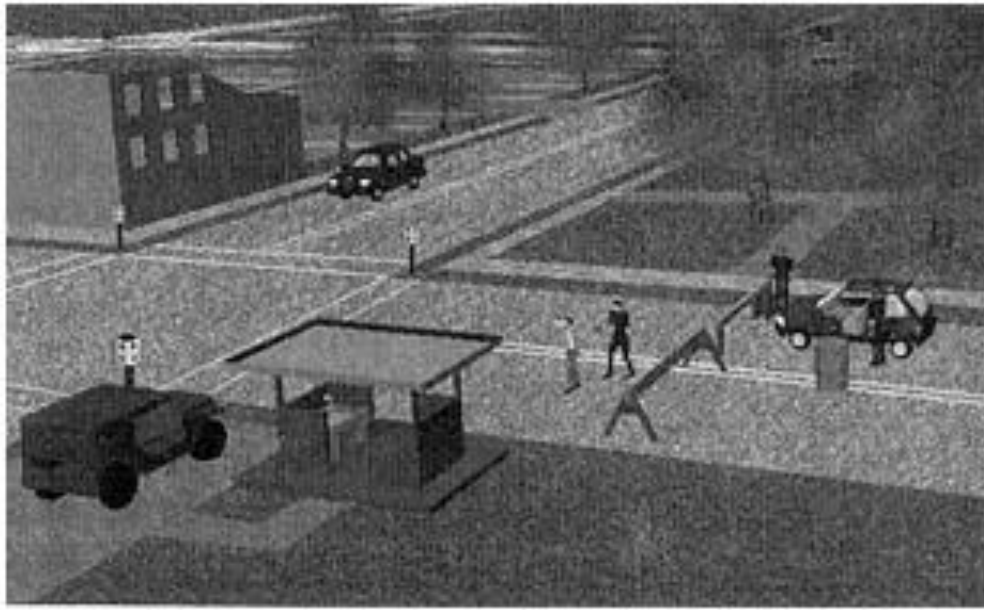


Figure 21: VET environment



Figure 22: Angry Facial Expression in VET

Embodied Characters

ECA's that are embodied with conversational models come to life to take on the role of the desired tasks that the designer creates. Many of them start with a simple wireframe, which outlines the basic shape of the character that is being portrayed. When these wireframes are embodied, there can be up to thirty three parameters of movements

in the jaw, lips, and teeth that can be manipulated (Massaro, 1977). The example shown below of an ECA has approximately 900 surfaces composed of polygons (Massarro, 1977).

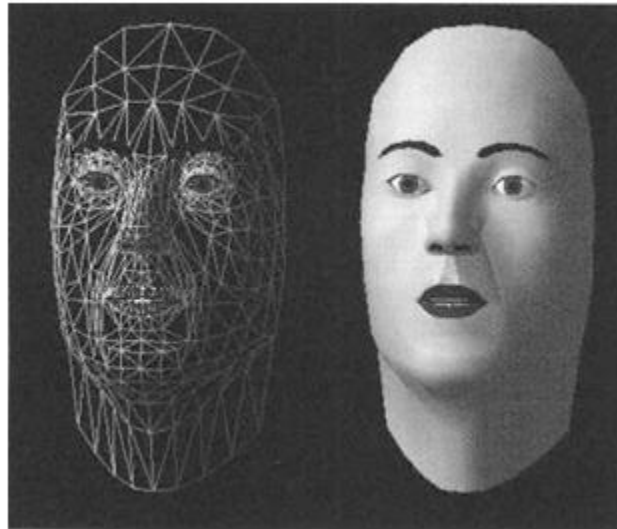


Figure 23: Polygon wireframe for an ECA



Figure 24: Syllable Pronunciations

Speech is another topic that is integrated into the design, and when considering speech, the inflection in the voice, pitch, and patterns play a significant role in how messages can be perceived.

The advancements in natural language processing have developed slowly, but the challenges still exist in fidelity, speech understanding, speech recognition, and lag in processing time of messages.

Evaluating Interactive Systems with Embodied Conversational Agents

Interactive systems and development of new technologies to leverage the technologies of the ECA's are important, but developers still must determine why and they are used (Sanders, 1998). Metrics should be put in place to track the progress and growth of these technologies (Sanders, 1998). One method is to compare a technique across other platforms and see if some of the needs overlap. Two things where this can apply to with Standardized Patient Training is the correctness of an SP evaluation form and the time to complete a training session. Testing and Evaluation has been also proven to be an effective method to document progress and growth (Sanders, 1998). Periodic reviews can also be held to document changes and action items. The end user drives the requirements for design, and the applications of systems engineering can assist in requirements development. Sometimes requirements can be ambiguous or out of scope, and these applications provide issues when considering the requirements for how and what would be desired to be tested. These user driven requirements can also affect the software that is created when the requirements change due to a change of high level objectives or realignment within the organization (Sanders, 1998).

Ethnicities in Embodied Conversational Agents

One of the benefits to standardized patient training with embodied conversational agents is being able to implement diversity. Diversity can be seen in ages, sexes, races, and geographic locations. Sexual orientation, religion, and many other distinguishing factors can cause an individual to be considered a minority in a particular social group. ECAs don't have an actual race, nor do computers. However, when ECAs can represent a particular ethnicity, a user can interact different based on the ethnicity of the agent and the user. Furthermore, when someone is part of the in-crowd, they are considered more socially attractive (Lee 1993; Stephan and Beane 1978).

Human Centered Distributed Conversational Modeling

Human Centered Distribution Conversational Modeling reduces the complex efforts of designing a conversational model with the assistance of an application called crowd sourcing (Rossen, 2009). The new knowledge that is created from crowdsourcing of the user's inputs provides an avenue to teach the VH a conversational model. One of the first steps in HCDCM is to collect a mass of stimuli to begin with. With this sample, a domain expert guesses what is said to the VH; and then later approximates the response of the VH. Later novices provide responses in the same fashion which gives new stimuli for the VHs. Later this is later peer reviewed by a subject matter expert and validates the use of it. Below is a diagram which was created by Brent Rossen which describes visually how HCDCM is designed. Later described below is Virtual People Factory, which allows this process to occur. The next image below is a diagram which shows the work flow of the process.

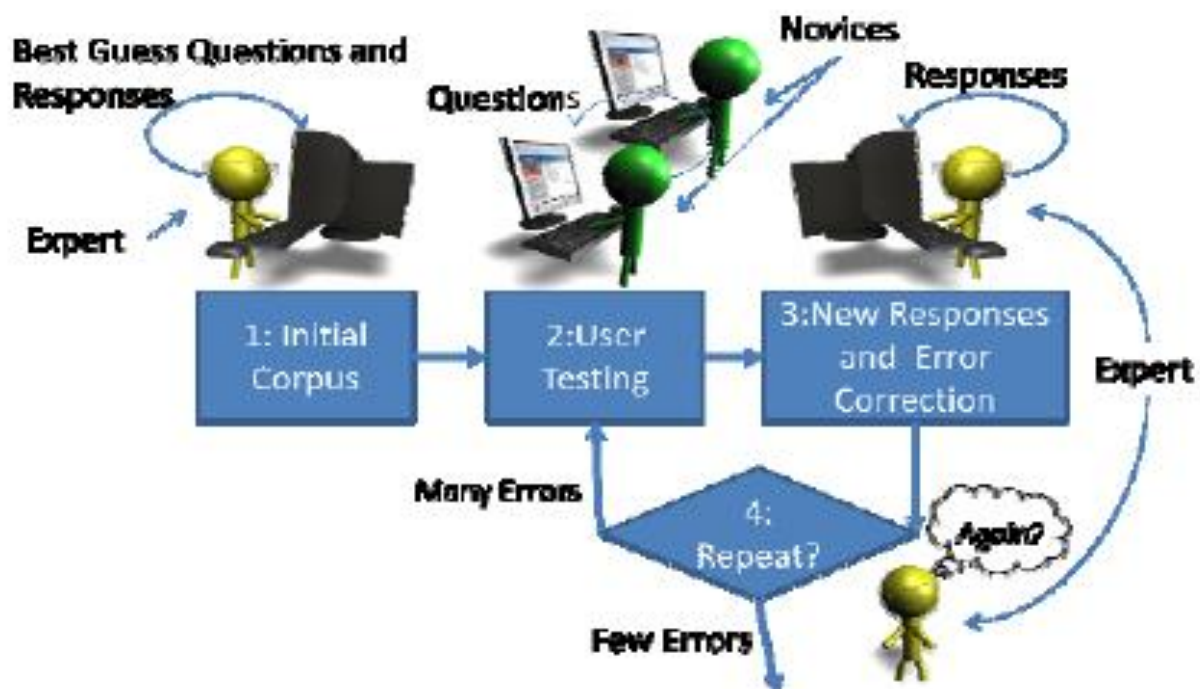


Figure 25: HCDCM Design

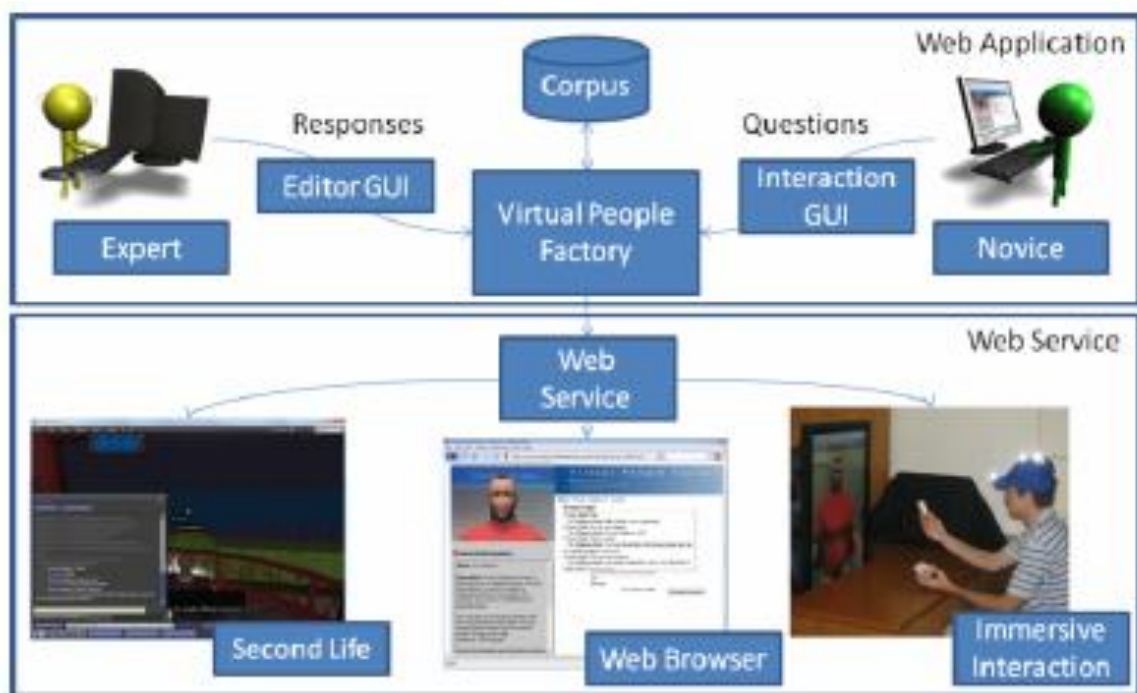


Figure 26: Virtual People Factory Workflow

A browser-based interview GUI allows an easy interface for communicating with the virtual human using a text to speech functionality. When matching items in the corpus, VPF uses a special approach that can recognize similar responses and elicit feedback accordingly (Rossen, 2009).

The Editor GUI is a domain where experts can create new questions and use VPF to interact with the use of phrases (Rossen, 2009). A web-based version of VPF allows the user to communicate with a patient with the use of an instant message based interaction. (Rossen, 2009). Simulated clinical interactions that rely more in voice activation than text to speech are applied with the use of an *Interpersonal Simulator* (Dickerson, 2005).

Bootstrapping the Creation of Virtual Humans from existing ones for SP Training

Virtual Humans can be applied to the training curriculums of several applications. Some popular areas where they can be useful are in military training, patient interpersonal

skills, and medical interpersonal training applications. Many training modules have had an interaction where the user interacts with a virtual patient and this practice session provides a great deal of part task training to master communication skills and bedside manner. Rossen's paper is quite interesting as it mentions the developmental efforts and time it would take to create new roles for each VH, and how this can be mitigated by using existing conversational logs from previous preceptors to assist in training SPs. This can provide a great training on teaching standardized patients how to rate. After an interaction is observed, they can complete a standardized patient checklist, and those results can be compared to that of a subject matter expert rater. This technique of using the utterances of a human from an interaction to create new VH that can take on the role of the Human is called bootstrapping (Rossen, 2010). A Role play trainer was created to assist in training standardized patients. The Roleplay trainer is still the web-based interface, but can be brought to life with integration of the interpersonal simulator. However, there is significant back-end coding that needs to be done to accommodate this. Below is a diagram that explains this logic.

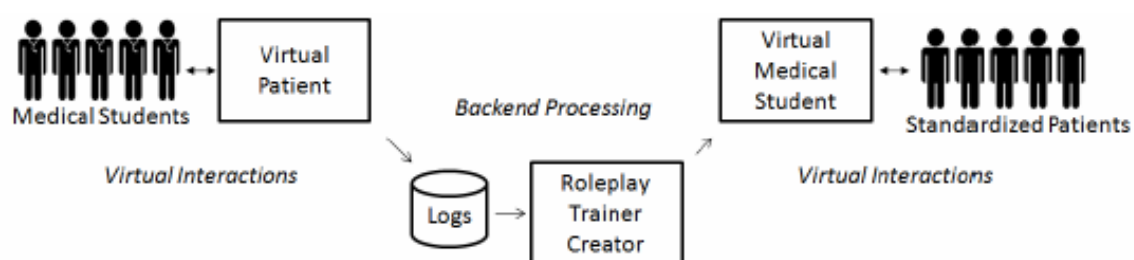


Figure 27: Virtual Medical Student Generation

One of the difficulties that the Role-play trainer faces is that when it creates a question asking VMS from the logs of the VP - Medical Student interactions, there are issues being able to have it say the proper questions in a logical order that flows similar to natural language. This developmental tool serves as a layer that works above the existing

Virtual People Factory web-site architecture (Brent, 2009). The Roleplay Trainer Creator is used to create models of conversation that works properly with the Virtual People Factory (Rossen, 2010).

A pilot study was performed to test how well the trainer worked on some users (Rossen, 2010). Surveys were used to get feedback on how well the pilot worked with users (Rossen, 2010). The purpose of this PhD dissertation is to expand on this study and strengthen the case that this training tool enhances proficiency for SPs in medical interviews. The pilot showed benefits in certain areas after the study (Rossen, 2010). One of the subjects that rated lower in his post survey might have potentially rated higher in his initial assumptions thinking he had a higher ability (Rossen, 2010).

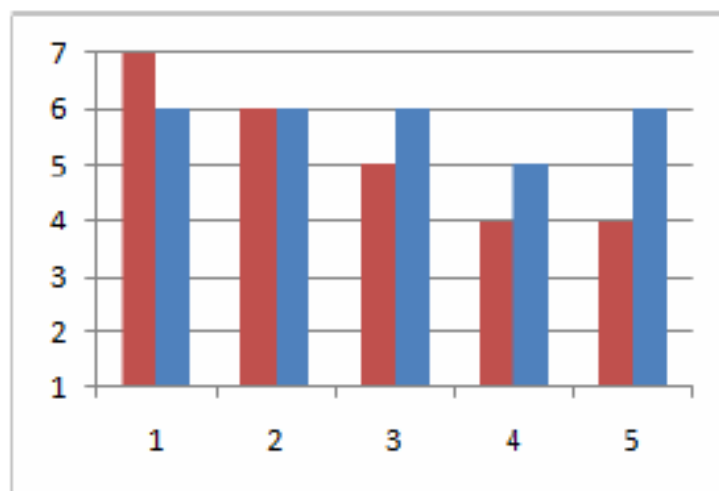


Figure 28: Comparing Preparedness between the Pre and Post Survey

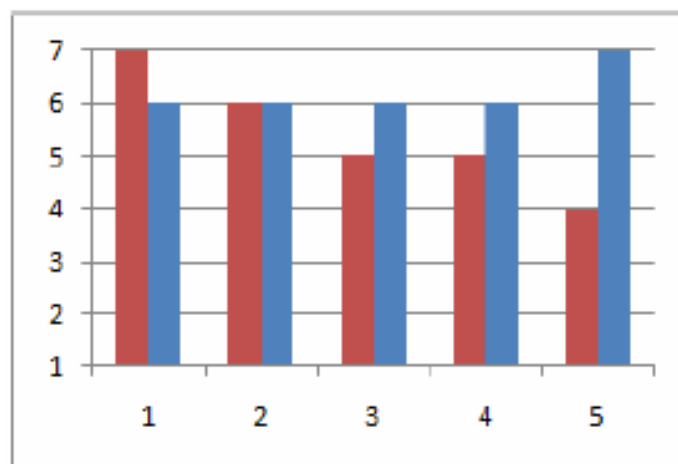


Figure 29: Comparing Confidence between the Pre and Post Survey

Additional applications that can be leveraged off the Role Play Trainer Creator is the ability for a VH instructors to perform debrief exercises where discussions and forums can be setup to review common areas for error along with opportunities for improvement and optimization (Rossen, 2010). In addition, inputs can be created for the VH to be more expressive and show more realism in interactions.

Participant Motives in Conversational Modeling

It was often believed that the developer was expected to create a system that made the user content, which is still true. However, now the user can also assist the developer with feedback for new creations in software development. Halan mentions that internal interactions can be applied to help in this user motivation. This same school of thought can be used as a proof of concept in SP training. SPs may have a hesitation in the credibility of a training system as it is a rather new technology that may have not built the desired reputation that matches up against a live role play training scenario with a medical student. However, their feedback can still be crucial in designing a system that is tailored to their learning curve and addresses the obstacles that are presenting in learning a task. Leaderboards, narratives, and deadlines have been used as great tools to gather attention and motivation; especially in social networks and media. However, the issue is that many of the medical professionals have very limited time (Halen, 2010). Below is an example of a leaderboard that was developed by Halen.

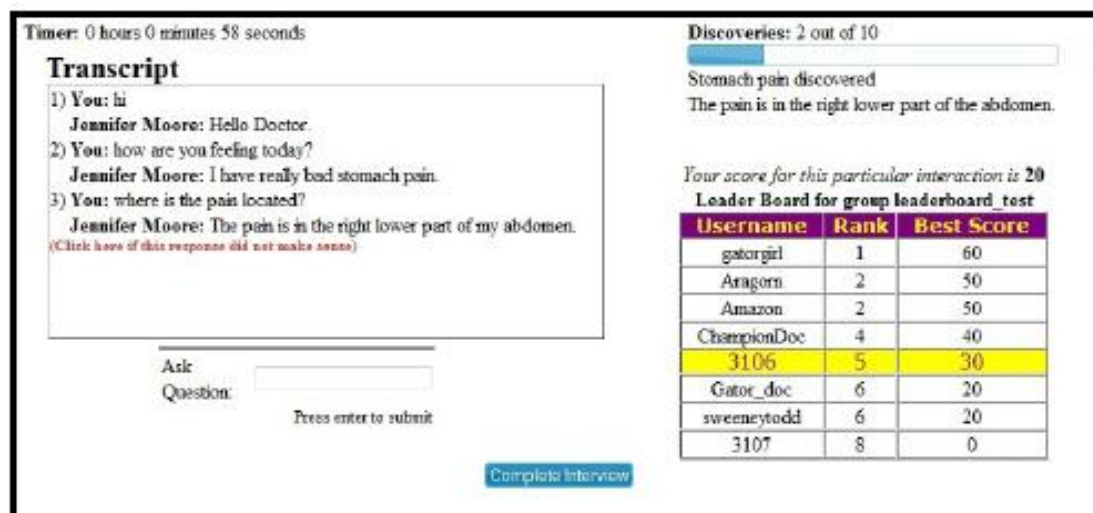


Figure 30: Leaderboard for Discoveries

Narratives were used to capture attention of the user and describe the patient's background information and were used as part of the user interest emails.

"Jennifer Moore, a 21-year-old female, presents to the emergency department complaining of lower right quadrant abdominal pain. . The onset of the pain was 1 p.m. The current time is 5 p.m. She has never had this kind of pain before and is extremely worried about what might be the cause."

Figure 31: Narrative

When a Deadline is pushed against the user, there is greater sense of urgency and motivation for the user to comply with the policy on time or earlier for attending the medical alert (Halen, 2010).

Virtual Patient Interactions

Foster's paper discusses possible integrations of Virtual Humans in psychiatry educational applications. Aside from Standardized Patient training, standardization can also be a challenge in other areas such as psychiatry. A Virtual patient provides assistance when there is a need to simulate an interaction. Below are some images of the interactions that were used for training.



Figure 32: Life-size Bipolar Virtual Patient

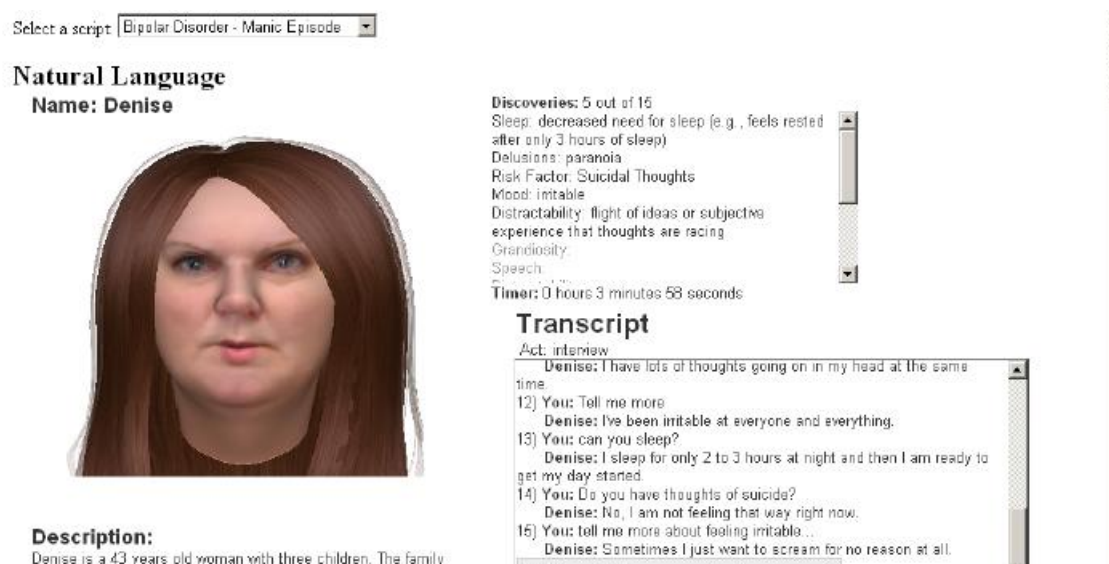


Figure 33: Web based Virtual Patient

One of the deficiencies in Standardized Patients is their memory of the case along with remembering the proper responses that are needed to provide a viable and realistic interaction. Some types of immersive training can be used to assist in memory. An application that can possibly be leveraged is a Virtual Reality Memory Training as explained in the article by Optale in 2009. This simulation assists in addressing memory

loss in the elderly (Optale, 2009). VR is an experiential interface where there is a sense of being at the place where the virtual reality has established (Optale, 2009).

When modeling Virtual Humans, being able to break down the process into Five Levels of Hierarchy allows the decomposition of a problem into subsystems so each portion can be approached one at a time and can later be integrated into a consolidated solution. This diagram can be seen below.

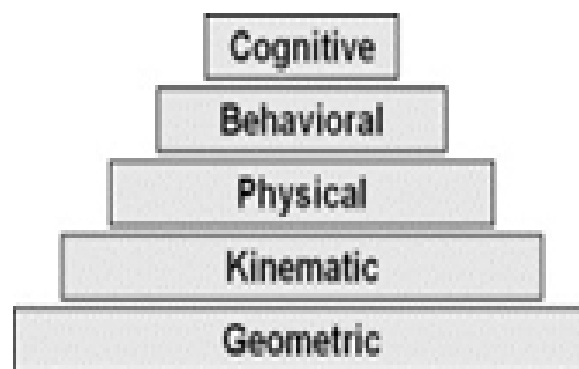


Figure 34: Five Levels of Hierarchy

Virtual Humans Providing Mentorship

Mentorship is a widely important goal among many organizations. The valuable lessons that leaders have acquired and passed on to new protégés, allows the protégés to become mentors as well. Embodied conversational agents can also become mentors (Sims, 2007). According to (Lester, 1997), motivation in people to learn can be enhanced. Virtual humans can provide support and motivation through the use of modeling and simulation. Role playing actors can be beneficial in standardized patient training as they can pose as a standardized medical student that interacts with a Standardized Patient who is in the process of learning a medical case. However, in order to perform as a successful mentor, they need to effectively model the behavior of a

successful mentor (Sims, 2007). Essentially, this type of smart avatar would be able to make decisions, inspire, and lead development as a real mentor would.

Mental Models and Immersion of the Future

If the time is truly taken out to understand what is distinct about humans, we can see that humans have the distinguishing characteristics to make themselves unique in several ways. These differences range from general ones such as gender and ethnicity to interesting ones such as types of technical advancement and the distinct way of wearing clothes. However, the true distinction that can be noticed by humans is the ability to create mental models. Ancient civilizations have implemented mental models that are similar to that of the modern day technologies that we have today for information technology, communication, and entertainment (Kurzweil, 2003). A very primitive form of virtual reality is the telephone. The telephone allows one person to communicate with another as if they were in the same room, when actually speaking from a distance (Kurzweil, 2003). Advancement in this area has now expanded to provide immersion in virtual worlds, and Kurzweil mentions that these virtual worlds are integrated into our contact lenses and woven into our clothing, so that the concept of computers as stand-alone systems are obsolete. Nanobots can also be used to intercept interneural signals to allow a human to enter a virtual world. Nanobots can also be used to perform surgical procedures to complete tasks while entering through natural orifices. It is also suggested by Kurzweil that by the year 2030, websites will be fully immersive experiences where emotions and experiences can be archived, and humans can increase their thinking with nanobot technology which are constantly enhanced with biological and nonbiological thinking. Kurzweil also notes that in the year 2030, we will not be able to tell the difference between biological people who have projected themselves into a fully immersive environment. One of the most crucial factors for success in virtual humans is

not only the level of complexity in technical advancement, but the realism and fidelity of the expressive characteristics. It's these characteristics that evolve into a personality.

Trusting Virtual Humans and Synthetic Thinking

Ethics and standards are policies that should be regulated as a safety precaution when it comes to relying more on conversationally modeled humans (Plantec, 2003). This can pose a risk as virtual humans are given more social responsibility and influence to the extent that it could possibly take over the boundaries of a live human (Plantec, 2003). The issues of the future that Plantec suggests are the increased dependence of virtual humans in the future as humans not only rely on them for completing tasks, but as companions, lovers, and confidants. There are also assumptions that humans will one day rely more on the company of a virtual human for comfort than a real human. When understanding consciousness and the quantum theory, Hameroff and Penrose suggest that the brain can process 10^{17} operations a second. Machines can process language at the unconscious level as much as we do. However, real organic brains use fuzzy logic which can accept inaccuracies and ambiguity; while computer brains are based on properly calculated algorithms based on rules which provide the illusion of a conscious personality (Plantec, 2003). However, a simulated organic brain that uses logic based on a fuzzy approach might provide a solution that would have some initial traits of a real brain.

Acting Concepts

One thing that is crucial to a standardized patient is having a firm grasp of acting concepts. Ed Hooks has 7 recommendations for concepts in face acting. They are as follows: The face expresses thoughts beneath, acting is reacting, know your character's objective, your character moves continuously from action to action, empathy is audience glue, and interaction requires negotiation.

Experiential Learning

Experiential learning is the act of being able to learn a task or a skillset as you go, and the experiences that you develop grow and mature the more active practice that you have. Experiential learning can be beneficial especially in standardized patient training as they the Standardized Patients can get up to speed on learning their lines for the SP case as they actively practice the case.

According to Baker, conversational learning provides a major training tool as learners can create new lessons learned as they experience new ones, and build new knowledge through their interactions.

Medical Focus Group

A Medical Expert Focus Group consisted of 3 Medical Doctors, a Director of SP Education at a College of Medicine, and two SP educators. The current way that SPs are trained at institutions is that they are emailed a medical case in advance. Upon receiving the medical case in advance, the SPs are expected to learn the case and then later have a discussion about it with the educators. After this, practice in role-playing is done, a dress rehearsal, and a dry run. Some of the less experienced SPs get more attention than the veteran SPs. An avatar that can bring an SP to proficiency would be an asset to the training along with providing the ability to have some extra training at home before the discussion with the educator.

A standard answer is expected from an SP, but often times the answer is not phrased exactly as scripted. To accommodate this, a guide is created to show acceptable answers that are within the acceptable range of answers. After this interaction is complete, a third person reviews the case and evaluates the communication skills. Something to consider is that medical students do not always ask questions in the same

way or order as scripted. An example of this is if the medical student were to ask the SP if they take drugs. The question would need to be executed to specify if it was referring to illicit drugs or prescriptions. Another example is if the Medical Doctor were to ask an SP “Do you read?”. The answer could either answer if the SP was able to read or be completely illiterate. The SPs should not be over trained to expect the questions in the same order as well to prevent them from being scripted. Both Avatars and Pre-recorded humans provide training value to SPs. The Avatar could be helpful to ask sensitive questions as some people feel more comfortable with an avatar as opposed to a human. However, this might pose a fidelity issue when having training time on a virtual human and then transferring to real humans. Although, SPs are now more tech savvy then before which could provide a more accepted response.

The drawbacks to having a Pre-recorded human are that the scripting is limited. If an avatar has an artificial intelligence system and natural language processing, it may be a better tool than Pre-recorded video delivery. The training effectiveness could be improved by examining how the SP performs during history taking. Behavioral responses are getting better now such as non-verbal cues, smiles, and nods that can be programmed. What would add to training effectiveness would be to have the avatars be plural instead of singular in a virtual world like second life. This virtual world could have a headshot of the individual who is being trained. This virtual world allows in giving context to the training. The benefits to having avatars are that certain parts can be exaggerated on certain factors to pick up details so that distracting components are eliminated.

One of the medical doctors mentioned that a comparison of the two depends on the case. In training medical students, a symptom which can’t be seen (headache, fever)

would be good with a Pre-recorded Human, but a sign which a doctors can see (trauma, swelling) would be best expressed with a Virtual Human. There was one case where an SP had a dislocated shoulder and the SP had to maintain a posture to simulate this. Something like this would have been a good approach with two solutions side by side. The benefits vary depending on the task. The benefits for each of the medical cases also vary depending on the learning style of the SP. Some SPs learn cognitively while others learn more visually or physically. If there is a high fidelity in robots, there is a large training value in integrating this technology in cross-training. One of the most important tasks in SP training is history taking. History taking probably amounts to about 30% of the interview, while the physical exam represents 40%. The remaining 30% involves advising of the patient, treatment, and counseling. At this point it cannot be predicted how SP training evolves, but it would be beneficial to also add trauma and post-op to some of the training of cases. In addition, the test, evaluation, and curriculum may have more stringent requirements in the future. Subject matter experts and healthcare professional should also be leveraged heavily to verify the validity of the cases and content. Both Avatars and pre-recorded humans are good tools, but together are powerful. One of the most important drivers is which provides the cheapest solution short-term and long-term.

Medical Student Feedback

Below is some open ended feedback that was given from some medical students that experienced an interaction with an SP, when taking their Step 2 CS board exam.

Medical Student#1: “The Step 2cs was pretty easy. The SPs didn’t really act the scenarios pretty well. I wish they had some type of simulation for me to practice doing interviews with different cases in a simulation for the board exam”

Medical Student#2: “The SPs acted very poorly when trying to portray a case, and I was just trying to get the exam over with. It’s kind of a hassle, because sometimes you have to go out of state to take the exam. I wish they had it via teleconference or could get credit through some kind of simulation as the test is pretty expensive as well”

Medical Student#3 “The SPs are horrible actors. When they cough, they don’t sound like they are really coughing, and the expression on their face doesn’t seem convincing enough to believe that they are really sick. Sometimes, they give the answer away if you don’t understand how they feel, which could give an unfair advantage”

Medical Student#4 “The SPs are very robotic, and they didn’t seem realistic at all in the case that they portrayed”

Medical Student#5 “The Step 2cs is not really a big deal if you pass, but it is a HUGE deal if you fail it, because it affects if you get into a good residency program. Sometimes the SPs give the answer away with the way they act, and if they feel a connection with the student, they might have a bias which could lead to an unfair advantage”

Medical Student #6 “My Step 2cs exam went fine. The SPs that I had were really good actors, and they portrayed the case very well. However, I did take the test in LA, and am not sure if the location plays a big role in the quality of SPs. Most people in the LA/Hollywood area are trained actors on the side. But at the same time, I was very nervous about the test and might have thought the SP was a good actor either way as I was just trying to pass the exam”

Research Statement

In the literature searches above, it can be seen that ECAs can provide a benefit when it comes to interfacing with virtual environments and acting as educational agents, mentors, and role players. According to Bailenson, humans feel more comfortable with avatar actors in embarrassing situations than they do with actors (J. N. Bailenson et al., 2005). In addition, some of the pressure is relieved when being cognizant of the fact that they are interfacing with an agent that is similar to a human-like character. This is one of the discoveries that we realize when dealing with some subjects who are interfacing with an avatar and some that are not.

The VHs provide a sense of detachment from real life which can increase the level in comfort although having a lower sense of presence (Conkey, 2010). This is important because prominent research has indicated that higher presence equates to more involvement and participation for a better training tool (Conkey, 2010). The study performed by Curtis Conkey which compared machinima training with video content delivery for soft skills training showed no change in the results. This means that the avatar based machinima had no difference in results than the training tool that had recorded human actors. What does this mean? It shows that a virtual human based tool for training applications that can provide a contribution to the medical community that is just as effective as a pre-recorded human version. Furthermore, this alternative provides more reuse, customization, briefing, debriefing, ethnical diversity in training, and updating capability. The objective of this research expedition is to see if change can be seen when applying these aspects to a Standardized Patient Training tool which compares Virtual Humans to Pre-Recorded Humans. Below is the list of assumptions that were gathered from the primary objectives:

Assumption 1:

Training SPs with experiential learning can reduce some time that SPs need to train on in a real simulated examining room for learning the task of knowledge transfer of a basic non-physical exam.

Assumption 2:

Using Virtual humans in training does not provide an immediate cost savings up front as there are detailed developmental efforts needed to create the layers for the intelligence that is needed for the ECAs. However, over time, there would be some savings in faculty time; resources, time to train, and possibly one day replace the need for live standardized patients in medical board exams.

Assumption 3:

Embodied conversational agents can simulate gestures, facial displays, and emotion that can provide communication skills training to help SPs (actors) possibly learn their lines/medical case/rating abilities in a shorter time frame. Thus, as mentioned, this could assist in reducing time and resources in the examining room.

CHAPTER THREE: RESEARCH METHOD

Scope

The scope of this PhD dissertation is “performance assessment”. This research method investigates if Virtual Human can perform just as well as a Pre-Recorded Human for an SP performance assessment. However, a follow-on study can utilize SPs in pilot studies for the prototyping and testing of a commercialized SP training solution in the future or using SMEs to code the rating ability of SPs. This educational module tests the knowledge of an SP case, but is not intended to replace a standardized patient training curriculum. It is intended to serve as an additional clinical skills tool which reduces the need for extended use on more expensive resources or expenses in staffing to educate patient interviewing skills, medical knowledge, and most importantly, remembering their lines. Furthermore, the focus is not to display a high fidelity human computer interface solution, or to use a heavy SLOC count to develop an original simulated solution for training, but to leverage existing technology to show that virtual humans can provide learning transfer that is similar to or better than traditional training with a medical expert , a pre-recorded human, reading scripts, or structured didactic training.

Test

The objective of the experiment is to investigate if there is a difference between the two groups (Virtual Humans and Prerecorded Humans) scores when participating in a performance assessment that tests their knowledge of an SP case. Subjects also complete surveys that analyze how useful and easy to use the study was, open ended feedback surveys, and the overall experience of the performance assessment. In short, do users treat avatars differently than humans in interactive computer based training? Research with 52 human subjects is necessary. Sufficient information could not be gained by other

research not involving human subjects. The population consists of UCF students (general UCF students, UCF College of Nursing, and volunteers from the UCF Medical College), NAWCTSD employees, and members of a UCF civilian organization called the LIFE Organization. The main criterion is that participants are comfortable with being audio recorded and familiar with using basic computer functions. Another requirement that was established was that at least half of the sample population must be above the age of 45 years of age. This sample population was justified by contacting several standardized patient educators from various medical institutions. When interviewing an SP educator at the University of New Mexico's Standardized Patient Program they said that they had SPs of all ranges from 18-80 years of age. However, the most commonly used were 50 and above. The Academic Program Manager for SP & Teaching Associate Programs at John Hopkins Hospital was also interviewed. This group hired 2 month old babies, 8 and 11 year olds, young teenagers for teenager cases, young couples who have a baby, but her average range was between the 45 and up range. Finally, the SP Educator at UCF College of Medicine said that the average age is above 50, but they are now having more people in the 30s come in now because of the economy and more inclined to seek work. As a result, the subjects were distributed in the mix as follows. Also below is the study design.

AVATAR
Ages 45 and up: 13
Ages 18-44: 13
TOTAL: 26 subjects

Pre-Recorded Human
Ages 45 and up: 13
Ages 18-44: 13
TOTAL: 26 subjects

Group TOTAL: 52

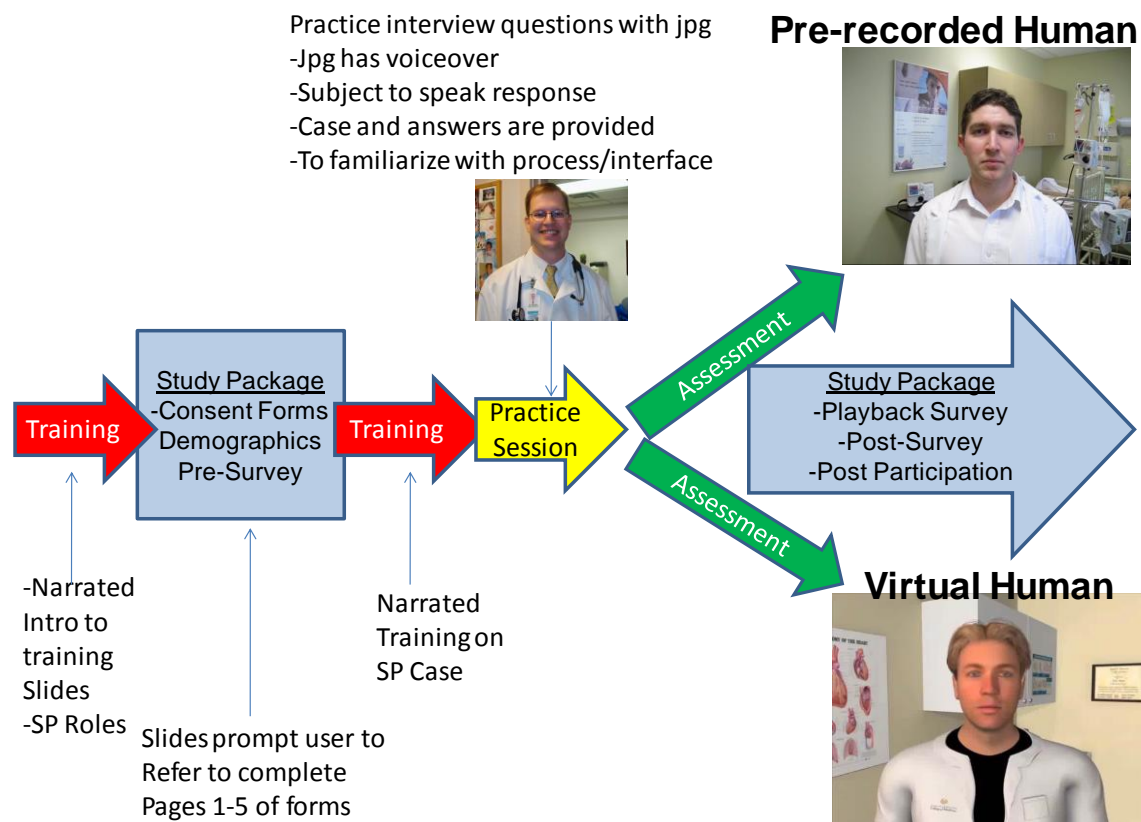


Figure 35: Study Design

Hypothesis

The hypothesis is that there is a difference between the two conditions (Virtual Human vs. Prerecorded Human) of testing the knowledge of a standardized patient case.

Manipulation

The only manipulation in the experiment was in the assessment. In reference to Conkey, 2010, a video training module was converted completely to avatars and the results showed that there was no change in results when comparing the two. The results were that it doesn't make a difference if you use video vs. avatars. Although that study used tasks in a hospital emergency room, the comparison seen in two training modalities is valuable. Preparing medical students to communicate with patients who

come from a variety of race, sexual orientation, financial status, age, and other considerations is very beneficial to medical students in assessing the patient in the most effective way

Summary of Manipulation

This study manipulated whether participants performed differently when faced with different modalities. This simulated a patient interview as participants were prompted to provide an answer upon each response of the conversationally modeled Virtual Human. The critical difference between the two conditions was the realism in the interaction. The avatars do not currently have the fidelity that compares to the realistic representation of a human image. This could affect the credibility that the participants are exuding, but the benefit is that this study provides the evidence that there are benefits, disadvantages, or no change based on the evaluation of the measurements provided

Power Analysis

The one independent variable is a performance assessment modality which is at two levels: Virtual Human and Prerecorded Human. Power must be collected after collecting data as it can't be controlled with these types of tests. The chances of having better power can be created by having larger samples or having data that has a farther margin apart. However, to get a second opinion, a power analysis was conducted to make an approximation on how many participants are needed to provide enough power to analyze the data to sufficiently correspond with the hypothesis by utilizing software for statistical analysis G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). This software allows the researcher to enter parameters of the study and then estimate the required sample size in order to detect a specified effect size. According to the use of G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), the medium effect size was at .04 with .80

power, which indicated a minimum n of 52. A similar study by Conkey (2009), estimated medium effect size in determining the number of participants that would be required. The participants are randomly divided into two groups for testing against the two major treatments to the training.

Recruitment

As mentioned, the target number of subjects that are needed is approximately 52. The recruitment occurs by using the NAWCTSD bulletin board, flyers, bulletins and emails to UCF groups. The UCF LIFE Organization is a committee of retired individuals that have several activities for participants to be involved in.

The complete study is done on a combination of a laptop, a PC microphone (as subjects are voice-recorded), and a paper packet that has paper-based measures. The Principal Investigator (PI) is standing by to provide additional assistance.

Study Workflow and Timeline

Upon the potential research subjects entering the study site, the PI provides each subject with an Informed Consent Document (ICD) and Privacy Act Statement. After informed consent has been administered and the ICDs have been signed by both the PI and the subjects, the PI provides a paper packet to each subject and introduces them to the computer-based training. The computer-based training directs them to the packets to complete. The packets include demographics and a pre-survey on their knowledge and confidence in standardized patients. After the forms are complete, the participants begin an instructional training session on standardized patient roles and a standardized patient medical case. This tutorial and case is based on the case that is present in the University of Florida's Virtual People Factory software. Within Virtual People Factory, a tool to train Standardized Patients was developed. Brent Rossen, the developer of this tool,

published a paper (referenced below) that discusses how the Roleplay trainer is used within VPF to be able to train SPs.

Brent Rossen, Juan Cendan, Benjamin Lok (2010). "Using Virtual Humans to Bootstrap the Creation of Other Virtual Humans". Short Paper. *10th International Conference on Intelligent Virtual Agents (IVA)*, Philadelphia, PA, Sept. 20-22, 2010.

The case was used from this web interaction link:

http://vpf.cise.ufl.edu/VirtualPeopleFactory/publicOverview.php?u=3535&script_id=424a

After the tutorial of SP roles and case is complete, a practice session occurs to allow the subject to be familiar with recording their voice, reviewing answers and the case to answer questions based on both, and navigating through the slides. The practice session asks 19 questions which add up to a total score of 38 (two points each). The same questions are asked later during the Performance Assessment where the answers and case is not available to review while in the section and the subjects have to come up with the answer based on what they learned in the Standardized Patient Training and practice session. After the practice session, half of the participants begin a performance assessment with a simulated virtual human (avatar), and the other half performs with a prerecorded human on video. Both modalities have the same content, resembling a standardized patient interview. The character on the screen (either an avatar or a pre-recorded video) asks questions regarding the case. The subject is to verbally respond as if they were truly conducting an interview interaction. As mentioned, this time the subject is not given the answers or the case as a reference. Instructions are given in the beginning of the session as well. This interaction is recorded. After the recording, the subject completes a performance assessment experience survey on the usefulness of interactive performance assessment and provides open-ended feedback about the study such as

strengths of the study, areas for improvement, and additional comments. Finally, a Post Survey with the same questions as the Pre-Survey and a Post Participation Form will be administered to the subject. The audio-recorded interactions of the subjects will be transcribed for single blinded grading. The scores of the population that used the avatar will be compared to that of the prerecorded human.

The timeline for the study

- 1-Consent forms/Privacy Act Statement: 7 minutes
- 2-Pre-survey: 3 minutes
- 3-Training (learning about SP roles and the Case): 5 minutes
- 4-Practice Session: 5 Minutes
- 5-Performance Assessment: 10 minutes
- 6- Perceived Ease of Use and Open Feedback Survey: 10 minutes
- 7-Post Survey: 2 minutes
- 8-Post Participation Information: 2 minutes

TOTAL: ~44 minutes

Required Equipment and Supplies:

- Laptop
- PC microphone
- Audio recording software
- Microsoft Office with PowerPoint (prefer 2007 and up)

The scores of the correct responses of the practice session will be graded using a ranking scale of 0-1-2. 0 is used to represent being incorrect, 1 for being partially correct, and 2 for being completely correct. This same scale for grading will also be used on the training after the responses are transcribed onto a sheet. The PI will be grading the assessments single blindly (all the data is transcribed onto a sheet, and is graded later blindly after all the data is collected without knowing which treatment each subject's data is categorized under. This is done by assigning each subject with an identification number. Funding was not available to hire other personnel to grade the papers.

These, along with the Perceived Usefulness, Open Ended Feedback Survey, and Pre and Post Surveys provide additional data that is represented from the study. To calculate results for these, we can use the T- test to see if there is a difference in the scores of the two groups.

Objective/Subjective Measurements:

Dependent variables will be the scores achieved on the interactions based on a ranking of 0-1-2. A Two sample T-test assuming unequal variance is conducted. All subjects have a potential of achieving a score of 38 (19 questions x 2 points each). An additional analysis technique can also be used to determine whether an effect seen in the data is due to the manipulation of the independent variable or the result of random variation.

Decision Rules

These rules will determine whether the H_0 will either be rejected or accepted. The two types of errors will be explained as follows:

Type I error = We reject H_0 when we should fail to reject $H_0 = \alpha = 5\%$

Type II error = We accept H_0 when it should be rejected = $\beta = 20\%$

=> Power: $1 - \beta = 80\%$

Justification for Exclusion of Specific Groups:

An age restriction was that the research subjects must be at least 18 years old.

Anyone with a history of seizures shall be excluded from participating in research involving the use of typical video and/or computer games. The PI, Joel Palathinkal, will have responsibility for administering informed consent, briefing research subjects, scheduling research subjects, collecting data, conducting analyses, writing report(s).

Janet Raskin, Research Assistant will assist in data collection.

Risks and Discomforts to Research Volunteers

Researchers conducting this study will be ready to destroy any subject's data forms at any stage of the study if there is a motive that displays that there are any health or fatality risks. During the study, all personal data or information (such as demographic data/performance data) will be secured under lock and key until destroyed in accordance with SECNAV M-5210.1 and as required by 32 CFR 219. Any subject identification keys used during the study will be kept separate from any signed ICDs and will be destroyed at the end of the study. This procedure will ensure that the participant's personal data cannot be used in any way that might impact his/her career, academic progress, or standing in their respective professional or educational communities. There is currently very minimal risk, and no opportunities for harm of research subjects is currently presented due to it simply being computer-based training that will require voice recording and surveys to be completed.

Maintenance of Experimental Data for Records.

The PI, Joel Palathinkal, will maintain all experimental data. Any data subject to the provisions of the Privacy Act will be stored under lock and key. Data will be received and procured via computer and audio recordings. The PI will be responsible for the data that is received and transmitted. The data will be transported via paper-based storage in a secure envelope, and electronic backups will be stored on a computer hard drive. Data will be managed by transcribing the interactions of the subject and training session onto a transcript. The transcripts will then be stored on a hard drive and each subject is assigned a special code number for blind grading after all the data is collected. A research assistant, Janet Raskin, will also assist in data collection but will provide all data to the PI for consolidation. The research protocol application package and document, with original signatures, will be stored in the files of the Committee for the

Protection of Human Subjects (CPHS) Administrator at NAWCTSD. A copy of the research protocol documentation and the recommendations of the CPHS will be forwarded to the Institutional Official (IO) for review and protocol approval. The PI will maintain the original signed ICDs under lock and key until the completion of data collection. Upon completion of data collection, all signed ICDs will be provided to the CPHS Administrator. The signed ICDs will be kept under lock and key in accordance with provisions of the Privacy Act. The signed ICDs will be secured under lock and key until destroyed in accordance with SECNAV M-5210.1. The Final Report/End of Project Report is due to the CPHS no more than two months after data analysis is complete. The CPHS considers research to be complete at the conclusion of the data analysis. DON HRPP considers a protocol active until data analysis is complete and therefore subject to Continuing Review.

CHAPTER FOUR: RESULTS

Demographics

The table below demonstrates the demographic pool of the subjects selected for the study. The age range was 19 to 68 with a mode of 22 (4 subjects out of 52 of age 22). A pilot study was conducted with 13 subjects to get familiarize with the data collection and overall study procedure. Subjects are very much evenly spread throughout the age. In Table 1, age distribution is categorized in three parts. The age group between 30 and 50 makes up for around 48% of the total subjects. The proportion between genders is not even but very comparable. The study is not inclined to any extreme side in gender. The Participants were predominantly Caucasian with almost 50% of the subjects followed by Hispanic of about 21%. The average assessment score seems to be spread evenly in all demographic categories. The only score that stands out is for Race-Asian. The mean assessment for Race-Asian is 30 with standard deviation of 6. This indicates the high variation among the Asian group.

Additional demographic data was collected and analyzed in the following table to establish each participant's education level and knowledge about the study. All subjects recorded their education before the study and this data is used to further analyze the mix among subjects. The data for education level is collected into two parts where education of Masters and PhD level is considered as Advanced degree. About 36% subjects have an Advanced degree and the mean assessment score is very comparable in this category.

Data is collected to capture proficiency levels in acting and healthcare experience. The subjects who responded with previous acting experience were around 27%. Similarly, subjects with healthcare experience were around 23%. The initial review of mean assessment scores between these categories indicates that their scores are very close and it seemed like their acting or healthcare experience did not skew scores.

Demographics Tables

Variable		Frequency	Percentage	Average	Std Dev.
Condition	Pre-recorded Human	26	50.0%	33.5	3.0
	Virtual Human	26	50.0%	33.9	3.9
Gender	Male	30	57.7%	32.9	3.8
	Female	22	42.3%	34.7	2.7
Race	Asian	6	11.5%	30.0	6.0
	African American	8	15.4%	32.3	3.3
	Caucasian	26	50.0%	34.2	2.5
	Hispanic	11	21.2%	35.0	2.2
	Native American	1	1.9%	38.0	0.0
Age	Below 30	15	28.8%	34.9	2.3
	Bet ⁿ 30 and 50	25	48.1%	33.2	4.1
	Above 50	12	23.1%	33.2	2.9

Table 2: Modality, Gender, Race, and Age Demographics

Variable		Frequency	Percentage	Average	Std Dev.
Advanced Degree	MS & above	19	36.5%	32.7	4.4
	BS & below	33	63.5%	34.2	2.7
Acting expe.	Yes	14	26.9%	34.6	2.2
	No	38	73.1%	33.3	3.8
Healthcare expe.	Yes	12	23.1%	33.7	2.6
	No	40	76.9%	33.7	3.7

Table 3: Education, Acting, and Healthcare Demographics

Open Ended Feedback Results

Participants to Virtual human training predominantly had following feedback:

- Slow down the voice (70%)
- Less engaging/reality as you know it is not human (65%)
- Easy to understand (30%)

Participants to Pre-recorded human training predominantly had following feedback:

- Easy to understand and work (75%)
- Little rushed into the interactive process (45%)

Hypothesis Testing

The question being asked is whether there is any difference in the performance assessment score between training methods of Virtual human and Pre-recorded human.

The six step hypothesis testing is conducted as below (Corty, 2006):

1) Test:

The question being asked is whether there is any difference in the performance assessment score between training methods of Virtual human and Pre-recorded human.

2) Assumptions:

- The samples of Virtual human and Pre-recorded human are independent as subjects are selected and assigned to training method randomly.
- Performance assessment score is a ratio level measurement (scores of 0 to 38)
- The Kolmogorov-Smirnov test of Normality is conducted on Assessment score to understand whether the data is normally distributed or no. The result is shown in below graph. As the p-value is less than significance level of 0.05, it can be concluded

that data follows normality. (If score does not follow normality then it can not conduct two sample t-test assuming unequal variances test)

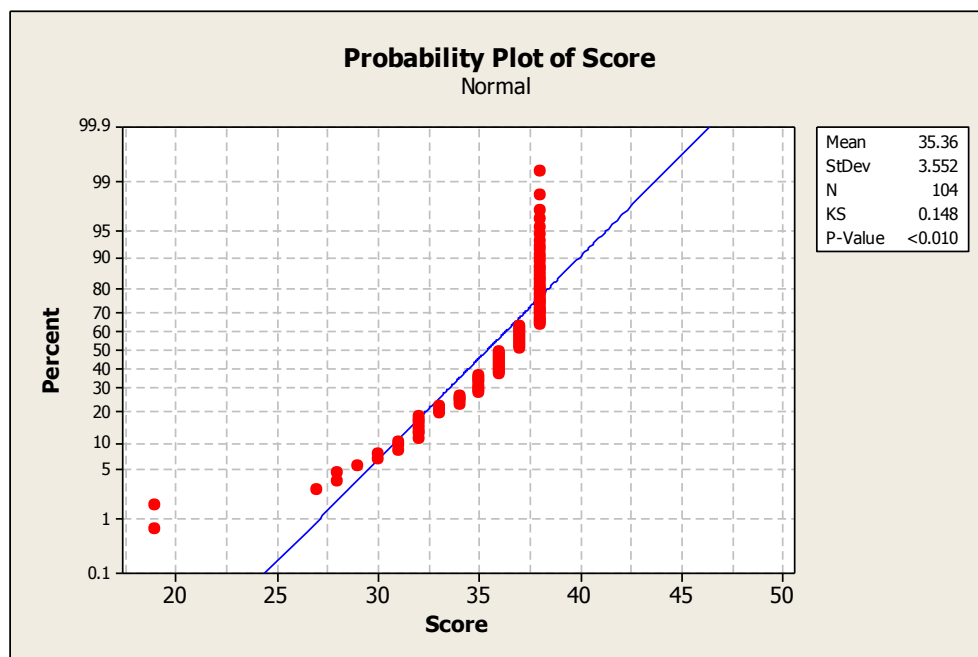


Table 4: Kolmogorov-Smirnov test of Normality

The partial descriptive statistics is listed below:

The standard deviation of both samples does not show big difference among them.

The larger value of std dev is not twice as large of the smallest std dev. Based on this, it can be concluded that homogeneity of variance is followed.

Training Method				
	N	Mean	StDev	SE Mean
Virtual human	26	34.23	4.03	0.77
Prerecorded human	26	33.46	2.98	0.59

Table 5: Training Method Homogeniety

3) Hypothesis:

There is a difference in two conditions (Virtual human and Pre-recorded human) when comparing the scores of a Performance Assessment in a Standardized Patient case

H_0 : Mean score of Virtual human = Mean score of Prerecorded human

H_1 : Mean score of Virtual human \neq Mean score of Pre-recorded human

4) Decision Rule:

We are willing to make 5% of Type I error so $\alpha = 0.05$. If p-value of the two sample t-test assuming unequal variances analysis is greater than 0.05 then that category is not significant and there mean difference is not significant, so we will fail to reject the null hypothesis. If p-value is less than 0.05 then the mean difference is significant and we will reject null hypothesis.

5) Calculations:

Microsoft Excel application is used for two sample t-test analysis, the output is shown below:

t-Test: Two-Sample Assuming Unequal Variances

	<i>Virtual Human</i>	<i>Pre-Recorded Human</i>
Mean	34.23	33.46
Variance	16.26	8.90
Observations	26	26
Hypothesized Mean Difference	0	
df	46	
t Stat	0.78	
P(T<=t) one-tail	0.22	
t Critical one-tail	1.68	
P(T<=t) two-tail	0.44	
t Critical two-tail	2.01	

Table 6: Results of a T-Test using Two Sample Assuming Unequal Variances

6) Conclusion:

Training method represents the data for Virtual human vs Pre-recorded human and Survey time represents the Practice session vs the performance assessment. The p-value for two samples of performances assessment scores is 0.44 which is greater than 0.05 → **Failed to Reject Null hypothesis** → We conclude that the Performance assessment method of Virtual human or Prerecorded human does not make significant difference on Assessment scores.

CHAPTER FIVE: DISCUSSION

Natural Language Processing

The advancements in natural language processing have developed slowly, but the challenges still exist in fidelity, speech understanding, speech recognition, and lag in processing time of messages. The Roleplay trainer could later be integrated with the Interpersonal Simulator to accommodate standardized patient training with artificial intelligence, voice recognition, and voice understanding.

Future Approaches

Coding for Interrater Reliability

An SP and SME could be required to complete a checklist to rate the medical student to teach medical student evaluation skills. The SP will complete this immediately after the role-playing exercise, and the SME will do this at a later convenient time. The role-playing performance exercise will be recorded, archived, and reviewed for coding by the SME. The SMEs will also be watching the archived video to rate how well the SP performed on the standardized patient encounter. In conclusion, the measures from the checklists will be a rater checklist from the SP and SME and the differences in results between the SP and SMEs in the checklists.

Using Real SPs

A study could be done with real standardized patient to show the learning transfer with real SPs so that a true application can be seen. Furthermore, if these studies prove to show benefit to the SPs, then the Roleplay Trainer could implement the Interpersonal Simulator to provide Conversationally Modeled Standardized Medical Students for SP training.

Automated Assessments

Finding ways for avatars to automatically assess the SP in conjunction to the training exercise can also provide value and reduction in unnecessary staffing.

Leveraging the Acting Community

In big cities like New York and Los Angeles, actors play significant roles like a woman named Quinn Lemly who plays a character that is staged in a 1940s theme where sparkling evening gowns are the center of attention (Lagnado, 2011). However, on some days that Quinn is not auditioning or performing, she will be on her way to the Weill Cornell Medical College to act as an SP (Lagnado, 2011). Some of these standardized patient opportunities can pay up to \$25 dollars an hour (Lagnado, 2011). The Weill Cornell Medical College in Manhattan is fully equipped with medical examining rooms that provide full capabilities for SPs to see where they made mistakes, see how it should be done, and provide two way mirrors for faculty to see the interactions (Lagnado, 2011).

With the current economy, the SP programs lead by Dr Kang in New York and Denise Lock from the University of Southern California's Keck's School of Medicine have seen a significant increase in actors looking for opportunities (Lagnado, 2011). Keck's medical program has hired individuals with acting experience in television shows and other commercials (Lagnado, 2011). Some portrayals that actors have performed are heart attacks, diabetes, Parkinsons, and hot flashes (Lagnado, 2011). The only issue is that sometimes the actors want to take the actors to an area that is out of scope of the scenario (Lagnado, 2011). However casting of these SPs can be difficult at times as Dr. Kang mentions "we want stars but we need to temper their star quality" (Lagnado, 2011). Some features that could be added in simulation to allow further applications to train acting would be to have more realistic facial expressions in avatars, and provide the ability to point to where the pain is.

Sixth Sense Technology

Gestures are used for every day emotions and expressions of different types of messages (Mistry, 2009). Some gestures are so common that they are assumed to have a meaning (Mistry, 2009). For example, a shrugging of shoulders might mean that a person is indifferent, or pointing at something might indicate a certain direction that must be approached (Mistry, 2009). These gestures can now be used as a human machine interface with the help of Pranav Mistry's Sixth Sense technology (Mistry, 2009). Such gestures like moving a page as done on the i-pad, pinching an image to expand its dimensions, or even joining two thumbs and index fingers while forming a rectangle to take a picture are effective in initiating commands (Mistry, 2009). This concept is to not limit the information gathering of the information age to only the mobile realm, but to leverage off of the mobile technology's and bandwidth to allow humans to interact with their biological surroundings, thus making any wall a touchscreen and making any piece of paper a tablet (Mistry, 2009). Thus, the digital world can be taken with the individual (Mistry, 2009). Using this technology into the immersive world assists in simulating a physical interview, where a standardized patient can be touched by a doctor for certain signs and symptoms.



Figure 36: Sixth Sense Technology

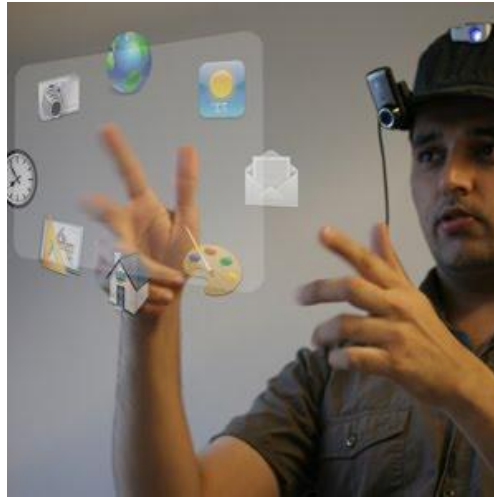


Figure 37: Sixth Sense Virtual Tablet PC



Figure 38: Sixth Sense Technology: Projection of Feeds

When seeing this technology, it can be observed that the technology that is currently being developed is not too far from what is seen in the movies (Mistry, 2009). A similar human machine interface was one that was portrayed in the movie “Minority Report”.

Holograms

Holograms are a popular technology that brings an added dimension to the realism that is desired in the entertainment industry. Now this technology is being leveraged for several military and medical applications. A briefing that was presented by M. Beth H. Pettit in the year 2009 to discuss the research plans for the joint medical simulation technology showed relevance to using holograms for training anatomy (2009, Pettit). This could be

useful in standardized patient education when dealing with cases that are related to post-operation. Later, a Request for Information (RFI) was issued by the United States Army that reached out to public companies and industry to produce a viable technical solution that provides three dimensional (3-D) holographic imaging (US Army, 2010).. These images are required to be functional without the need of 3-D glasses (US Army, 2010).. The overall mission and desires are to improve the quality of training that exists and allow better technology to bring troops to a higher level of competency in medical specialties (US Army, 2010).

Stretching Visual Images to Reduce Pain

A psychological test was completed to see if the stretching of visual images of a body could affect tolerance to certain types of pain (Mancini, 2011). When a subject was given an amount of heat to their hand, the pain was tolerable when the hand was seen on a reflection where the hand was stretched to seem larger than it was (Mancini, 2011). Two lessons can be gathered from this. One is that when we see some body parts being stretched, it allows in some areas to be more tolerable to pain (Mancini, 2011). Second, when the hands were out of sight, the pain was less tolerable, so it can be concluded that the hands must be seen to endure a larger tolerance.

Similarity Effects to Teaching

In a study that was performed by some researchers, a total of 257 subjects were involved in a study that involved a training in excel, where different questions were asked by an avatar (Behrend, 2011). The similar looks of the trainer had less of an impact than the subjective inclination to the educator (Behrend, 2011). The subjective feedback was compared with the objective feedback and this study proves that projections on how the training will perform could be better visualized when utilizing a subjective format (Behrend, 2011). Subjects provided inputs on how much they liked the trainer (Behrend, 2011). The

avatar seemed to be most engaging when it had similar characteristics and representations of the user, and possessed similar teaching styles (Behrend, 2011). When training standardized patients, the applications of avatars that have a closer relevance to the user might provide a more powerful training tool.

Using Artificial Intelligence for Resolving Disputes

According to an article in the Wall Street Journal, GE is using a technology to handle resolutions that are relevant to disputes in the oil and gas sectors (O'Connell, 2011). This software handles the bidding process and works to come to a settlement (O'Connell, 2011). The software is managed by an engineer, and will provide cost savings in legal fees and personnel (O'Connell, 2011). Although this system was criticized for not having a human element, if effective it could be beneficial to the energy arena (O'Connell, 2011). Similarly in medical simulation, if avatars prove to be just as effective as they were in this study, there is potential for cost savings and process optimization when viable logic in medical decisions are displayed.

Publication Status

This dissertation Abstract has been submitted to IITSEC (simulation conference), and will be re-submitted as a short paper that summarizes the study and objectives. This dissertation was also accepted as a Poster at the Society of Simulation 2011 conference, but will be expanded on and re-submitted as a paper as well.

Conclusion

After the analysis was complete, it was seen that there is no difference in standardized patient performance when comparing a Virtual Humans to a Pre-Recorded Human. Although there is no change, this still shows value when using simulation. When the use of the Virtual Human simulation provides the similar results as the pre-recorded human, the Virtual Human

is a better solution as it is more customizable, allows diversity, allows synthetic voice for different learning objectives, and is on demand. In addition, almost all aspects from the PANAS scale showed positive results especially in Interest, Enthusiasm, and Proud.

Discoveries

Some discoveries that were noticed were:

- Some subjects prefer using an Avatar because they feel more at ease
- Some subjects don't like the Avatar, because it is not as real
- Younger people like to use the Avatar because it is more similar to video games and is "cooler"
- One subject said that he didn't even focus on the avatar, because he was too worried about remembering his lines. He had a picture of the lines in his head instead.
- One subject said that when he looks at a character, he focuses on the background and if something in the background doesn't relate to the character he loses his focus
- Some subjects reach for their back or stomach in real life when speaking to the avatar/pre-recorded human.
 - Some said this was because it helped them remember
 - Some said that it was because for a moment, they thought they were talking to it
 - Some said it was to be in "character"

Below is a table that calculates the words per minute when comparing the two modalities. On the table, it can be noticed that the Pre-recorded Human has a higher words per minute count than the Virtual Human. The Virtual Human was purposely created to have a more synthetic sound compared to the Pre-recorded Human. The synthetic voice is good for testing the technology, but a more human-like voice is beneficial for high fidelity clinical skills training. Some examples of software that can be used are NextUp Talker, Cepstral, and AT&T Natural Voices.

Table 7: Virtual Human vs Pre-Recorded words per minute comparison

Question #	Virtual Human Time in seconds	PreRecorded Human Time in seconds	Words	Virtual Human time in minutes	Pre-Recorded Human time in minutes
1	1	1	6	0.02	0.02
2	2	1	6	0.03	0.02
3	2	2	7	0.03	0.03
4	3	2	9	0.05	0.03
5	2	1	5	0.03	0.02
6	3	3	10	0.05	0.05
7	5	4	18	0.08	0.07
8	3	3	7	0.05	0.05
9	3	3	7	0.05	0.05
10	4	3	10	0.07	0.05
11	3	2	11	0.05	0.03
12	2	2	6	0.03	0.03
13	3	3	7	0.05	0.05
14	3	2	9	0.05	0.03
15	1	1	3	0.02	0.02
16	2	2	5	0.03	0.03
17	3	2	8	0.05	0.03
18	2	2	6	0.03	0.03
19	2	2	8	0.03	0.03
Pauses between phrases	9.5	9.5	0	0.16	0.16
TOTAL	58.5	50.5	148	0.98	0.84
				Virtual Human words per minute	Pre-Recorded Human words per minute
				148/.98	148/.84
				151.7948718	175.8415842
				152 wpm	176wpm

APPENDIX A: IRB APPROVAL FORMS



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: UCF Institutional Review Board #1
FWA00000351, IRB00001138

To: Joel J. Palathinkal

Date: February 14, 2011

Dear Researcher:

On 2/14/2011, the IRB approved the following minor modification to human participant research until 02/01/2012 inclusive:

Type of Review: IRB Addendum and Modification Request Form
Modification Type: An additional Informed Consent has been approved for use. If study is funded, participants will receive monetary compensation. If study is not funded, there will be no compensation. PI will use appropriate consent document with participants.
Project Title: Virtual Human vs Video Recorded Training in Standardized Patient Interviews
Investigator: Joel J Palathinkal
IRB Number: SBE-11-07422
Funding Agency: Naval Air Warfare Center Training Systems Division(NAWCTSD)
Grant Title:
Research ID: NA

The Continuing Review Application must be submitted 30 days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form cannot be used to extend the approval period of a study. All forms may be completed and submitted online at <https://iris.research.ucf.edu>.

If continuing review approval is not granted before the expiration date of 02/01/2012, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Joseph Bielitzki, DVM, UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 02/14/2011 12:26:14 PM EST

IRB Coordinator



DEPARTMENT OF THE NAVY
NAVAL AIR WARFARE CENTER TRAINING SYSTEMS DIVISION
12350 RESEARCH PARKWAY
ORLANDO, FLORIDA 32826-3275

IN REPLY REFER TO
3900
Ser HRPP/192
MAY 16 2011

From: Commanding Officer, Naval Air Warfare Center Training
Systems Division
To: Mr. Joel Palathinkal

Subj: APPROVAL CERTIFICATION FOR PROTOCOL, TSD 192, VIRTUAL
HUMANS VS PRERECORDED HUMANS IN STANDARDIZED PATIENT
TRAINING

Ref: (a) 32 Code of Federal Regulation 219

Encl: (1) Informed Consent Document and Privacy Act Statement

1. This protocol is approved under the authority of Department of Defense Navy Assurance Number DoD N-40037 and in accordance with reference (a).

2. Exempt reviews of Protocol TSD 192, Virtual Humans Vs Prerecorded Humans in Standardized Patient Training, were conducted by the Naval Air Warfare Center Training Systems Division (NAWCTSD) Committee for the Protection of Human Subjects (CPHS) Chair on 6 May 2011 and Vice chair on 3 May 2011. It was determined that the risk of the study is minimal and that the study qualified for an exemption under Category 1. The CPHS Chair and Vice Chair recommend approval of the protocol as written.

3. Your protocol expires on 2 May 2012, or 364 days from the first CPHS review date. Should there be a need to extend this protocol beyond the one year approval period, you are responsible for submitting an Application for Continuing Review (CR) and supporting documentation to the NAWCTSD CPHS in sufficient time to allow for appropriate review and re-approval before the end of the current approval period. Human subject research shall not be conducted outside of an approval period. If the approval period expires, Federal law requires that the CPHS temporarily suspend the study and that work involving human subjects temporarily cease. CR must be properly completed and re-approval granted before the end of the current approval period in order to avoid interruption of the research.

4. It is your responsibility, as the Principal Investigator, to ensure that the protocol is being followed as planned and any adverse events or unanticipated problems are reported to the

Subj: APPROVAL CERTIFICATION FOR PROTOCOL, TSD 192, VIRTUAL
HUMANS VS PRERECORDED HUMANS IN STANDARDIZED PATIENT
TRAINING

CPHS. Should there be a need for modifications to the approved protocol, you will be required to submit an Application for an Amendment to the CPHS for review and approval prior to implementation.

5. Please provide a copy of enclosure (1) to each voluntary subject prior to their participation in the study.

6. A Final Report is due to the NAWCTSD CPHS no more than two months after data analysis is complete. The CPHS considers a research study to be active until data analysis is complete and subject to CR.

7. Should you have any questions, please contact the NAWCTSD CPHS Administrator at (407) 380-4320.



W. H. REUTER

APPENDIX B: STUDY PACKET

APPROVED MAY 16 2011. **This document may NOT be used after** MAY 02 2012

APPENDIX A.
INFORMED CONSENT DOCUMENT
(EXEMPT STUDY)

1. You are being asked to voluntarily participate in a research study entitled, Virtual Humans vs. Prerecorded Humans in Standardized Performance Assessment. You will be asked to participate in an interactive training session and complete some surveys regarding your training experience. You must be 18 years old to participate in this study. The researchers expect that there will be approximately 52 research subjects participating in this study. It is expected that this study will take 45 minutes. Breaks will not be scheduled during this study; however, you may take breaks as needed.

The interactions from this assessment will be audio-recorded. Your name will not be used during the training session. Please select one of the following options:

I consent to audio recording: Yes _____ No _____

2. There are no expected risks for taking part in this study. Information you supply is not traced back to your name. Additionally, there are no known or expected risks to a pregnant woman, the embryo or fetus.

3. Please understand that for this study, you will receive no direct benefit other than the knowledge that your participation may aid efforts to improve the level of quality of standardized patients.

4. Your confidentiality during the study will be ensured by assigning you a coded identification number. Your name will not be directly associated with any data. Any subject identification keys will be destroyed at the end of the study. This procedure will insure that your personal data cannot be used in any way that might impact your career, academic progress, or standing in your respective professional or educational communities. Please understand that all personal data or information (such as demographic data/performance data) will be secured under lock and key until destroyed in accordance with SECNAV M-5210.1 and as required by 32 CFR 219.111(a)(7). Your signed Informed Consent Document will be kept under lock and key separately from the subject identification key during the course of the study.

5. Should you have questions concerning the research described in this Informed Consent Document or questions concerning research-related injury, please contact the Principal Investigator listed below. Additionally, if you so desire, you may contact the Principal Investigator for a copy of any publication resulting from this study. Note that after the practice session and performance assessment portions are complete, the audio where you performed an interaction will be played back to you upon request, and then you will later be asked to complete a Perceived Usefulness and Open ended feedback survey. No results will be shared with other parties. If you have questions concerning your rights as a research subject, please contact the CPHS Chair or Vice Chair (see contact info provided below).

APPROVED MAY 16 2011 **This document may NOT be used after** MAY 02 2012

- a. Principal Investigator: Joel Palathinkal
Activity: NAWCTSD Orlando
Mailing Address: 12350 Research Parkway, Orlando, FL 32826-3275
Code: 4.6.1.5 Phone: (407) 380-4128 E-mail: joel.palathinkal@navy.mil
- b. Project Manager: Kristi Wiegman
Activity: Section 219 Funding NISE-WDP
Mailing Address: 48150 Shaw Road, Building 2109
Paxtuxent River MD, 20670
Code: 4.0T Phone: (301) 342-0168 E-mail: Kristi.Wiegman@navy.mil
- c. CPHS Chairman: Dr. Katrina Ricci
Activity: Naval Air Warfare Center Training Systems Division
Mailing Address: 12350 Research Parkway, Orlando, FL 32826-3275
Code: 4.6.5.1 Phone: (407) 380-4661 E-mail: Katrina.Ricci@navy.mil

- 6. Participation in this research study is voluntary. You may choose not to participate. If you decide to withdraw from further participation in this study, there will be no penalties. To ensure your safe and orderly withdrawal from the study, please inform the Principal Investigator listed in paragraph 5.a. of this document.
- 7. Your participation in this study may be stopped by the investigator at any time without your consent if it is believed the decision is in your best interest. There will be no penalty or loss of benefits to which you are otherwise entitled at the time your participation is stopped.
- 8. No out of pocket costs to you may result from your voluntary participation in this study.
- 9. Official government agencies may have a need to inspect the research records from this study, including your records, in order to fulfill their responsibilities.
- 10. A Privacy Act Statement concerning this research protocol is attached as Appendix B and will be provided to you prior to your participation in this study.
- 11. It is important that you understand what has been explained in this Informed Consent Document about your participation in this study. If you have any questions or concerns about this study and its related procedures and any risks that may be associated with your participation, please talk with the Principal Investigator listed in paragraph 5.a. All of your questions should be answered to your satisfaction.

TSD NUMBER: NAWCTSD.2011.0008, TSD 192

Protocol Title: Virtual Humans vs Prerecorded Humans in Standardized Patient Training

APPROVED MAY 16 2011. **This document may NOT be used after** MAY 02 2012.

12. By your signature below, you give your voluntary informed consent to participate in the research as it has been explained to you, and you acknowledge receipt of a copy of this document for your own personal records.

_____	_____	_____
Volunteer Signature	Printed Name	Date

_____	_____	_____
Investigator Signature	Printed Name	Date

PRIVACY ACT STATEMENT

1. Authority. 5 U.S.C. 301
2. Purpose. The purpose is to investigate if training with a virtual human will provide similar or better results than with a prerecorded human.
3. Information will be collected in an experimental research project entitled Standardized Patient Training to investigate if training with a virtual human will provide similar or better results than with a prerecorded human.
3. Routine Uses. The data collected will be used for analyses and reports by the Departments of the Navy and Defense, other U.S. Government agencies, and authorized government contractors. Additional use of the information may be granted to non-Government agencies or individuals by the Navy Surgeon General following the provisions of the Freedom of Information Act or contracts and agreements. I voluntarily agree to its disclosure to the agencies or individuals identified above, and I have been informed that failure to agree to this disclosure may make the research less useful.
4. Voluntary Disclosure. Provision of information is voluntary. Failure to provide the requested information may result in failure to be accepted as a research volunteer in Standardized Patient Training.
5. Appendix A. Appendix A of this protocol package contains the [*Informed Consent Document or Protocol Information Document*] for this study. A copy of the approved/dated [*Informed Consent Document or Protocol Information Document*] is to be provided to each research volunteer prior to the start of their participation in this study.

Demographics Form

Date:

Participant ID:

Demographic Info

1) Gender:

2) Age:

3) Race: other:

Please select all that apply:

Asian

☐

Caucasian

☐

Black

☐

Hispanic

☐

4) Highest level of education (including your current class level, if applicable – e.g., freshman, sophomore, junior, senior, graduate student):

5) Do you have any acting experience?

☐

Yes

☐

No

If yes, where you have you had acting experience

6) Do you have any experience in healthcare?

☐

Yes

☐

No

If yes, please give the following for each healthcare job you have held:

Job title:

Employer:

Years Experience:

Job title:

Employer:

Years Experience:

7) Are you bilingual?

If yes, what languages are you fluent in?

Pre Survey

POSITIVE AND NEGATIVE AFFECT SCHEDULE (PANAS)

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now, that is, at the present moment. Use the following scale to record your answers.

1	2	3	4	5	6	7.
Not at all	close to nothing	a little	moderately	moderately more	a lot more	extremely
_____	interested			_____	irritable	
_____	distressed			_____	alert	
_____	excited			_____	ashamed	
_____	upset			_____	inspired	
_____	strong			_____	nervous	
_____	guilty			_____	determined	
_____	scared			_____	attentive	
_____	hostile			_____	jittery	
_____	enthusiastic			_____	active	
_____	proud			_____	afraid	

For the following please fill a number on a scale of 1-7 as appropriate in the blanks for each question. Leave the answer blank if you don't understand the question. The survey questions are taken from Davis' perceived usefulness, perceived ease of use, and user acceptance of Information Technology.

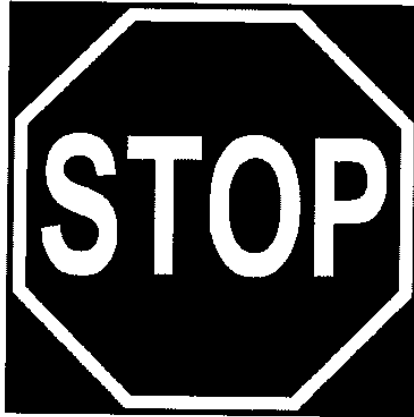
Strongly Disagree	Somewhat Disagree	Disagree	Neutral	Agree	Somewhat Agree	Strongly Agreed
1-----	2-----	3-----	4-----	5-----	6-----	7-----

Perceived Usefulness

- __ Training would be difficult to complete without interactive computer based training
- __ Using interactive computer based training gives me more control of my training
- __ Using interactive computer based training improves my performance
- __ Using interactive computer based training addresses my training needs
- __ Using interactive computer based training saves me time
- __ Using interactive computer based training allows me to complete my tasks more quickly
- __ Using interactive computer based training enhances my effectiveness
- __ Using interactive computer based training makes it easier to learn
- __ Overall, I find interactive computer based training useful in learning

Protocol Number: NAWCTSD.2011.0008, TSD 192

Protocol Title: Virtual Humans vs. Prerecorded Humans in Standardized Patient Performance Assessment



Please do not continue until guided by instructor

Protocol Number: NAWCTSD.2011.0008, TSD 192

Protocol Title: Virtual Humans vs. Prerecorded Humans in Standardized Patient Performance Assessment

Standardized Patient Case and Answers for Training Interview

Description: Instructions

In this interaction you will be talking with a Virtual Medical Student. The student will ask you a series of questions to conduct a patient interview. Please respond to the Virtual Medical Student as you would a real medical student. You respond by speaking your response as if you were conducting an interview. Please read through these materials as you will be able to reference them during your interaction.

Scenario

You are being seen in clinic because you have been having pain in the middle of your stomach for the past two months. You have also been having low back pain for two or three months.

You have been taking a Zestril once a day for your high blood pressure for the past 4 years. Your blood pressure has been well controlled on the medication. You take the medicine consistently. You do not have any symptoms associated with your high blood pressure. In fact, it was discovered on an annual physical. High blood pressure runs in your family. You come to the clinic twice a year for your prescription and blood pressure checks. You also check your blood pressure at Publix and Eckerd's from time to time and it is always in the normal range.

Your back pain started two to three months ago (you're not sure) after you injured it at work while lifting boxes. You work in the mail room of the post office. The back pain is located in the small of your back. The pain is a constant, dull aching located deep in the center of your back. You have been able to work but the pain gets worse as the day progresses so that by mid afternoon you can't sit comfortably. You have curtailed your activities and taken aspirin. You take adult strength aspirin, two pills, two to three times a day. The aspirin gives you a little relief from the back pain, but the back pain still bothers you. Sitting for prolonged periods and bending over make the pain worse. Lying down is the most comfortable position for relief. You tend to shift your weight to the right side when sitting and standing. On a scale of 1-10, with 10 being the highest, you would rate your pain as 8.

You have been taking aspirin for as long as your back has been hurting. Your stomach has been hurting for about the same amount of time, maybe a little less time. Sometimes the stomach pain feels better after you eat.

Past Medical History

You have generally been in good health. You have had no surgeries, blood Transfusions, serious illnesses or injuries.

Hospitalizations: No hospitalizations.

Allergies: Penicillin – causes a rash, Shellfish - swollen lips

Chronic Illnesses: High blood pressure. Diagnosed 4 years ago.

Protocol Number: NAWCTSD.2011.0008, TSD 192

Protocol Title: Virtual Humans vs. Prerecorded Humans in Standardized Patient Performance Assessment

Immunizations: up to date

Medications: Zestril 1 per day.

Over the Counter Meds: Aspirin (non enteric coated) for back spasms, tums for stomach ache

Illicit drugs: none

Social History

You are married and in a monogamous relationship.

Education: BS in Business

Employment: Post Office

Exercise: You walk for about 30 minutes a day

Tobacco: You smoke a half pack of cigarettes a day, and you started smoking when you were 18

Alcohol: 2-3 bottles of beer a week

Diet: You eat a quick breakfast, occasional fast food for lunch, and usually have dinner at home. Overall, you eat well.

Sexual History

You have been married for 10 years. You are sexually monogamous. You have never had a sexually transmitted disease.

Family History

Both your parents are gone, your Dad died a couple years ago, and your Mom died when you were twenty

Mother had a heart attack

Father died of colon cancer

Presentation/ Emotional Tone

You are concerned about the back problem and stomach problems because they are not resolving on their own. You are particularly concerned because your father died of colon cancer which presented first as stomach problems. You will answer questions cooperatively regardless of the interview style. You are neatly groomed, dressed in pants and a nice shirt.

Opening Statement

"I've been having this awful stomach pain, and it's starting to worry me."

Review Checklist

1. I take a Zestril once a day to control my blood pressure.
2. My blood pressure was diagnosed 4 years ago.
3. I have been having awful stomach pain
4. The stomach pain is gnawing, and sometimes it's sharp, especially at night
5. I do not have any symptoms related to my blood pressure (headache, chest pain)
6. My back pain started two or three months ago.
7. The pain is constant, deep, and aching.
8. The pain is in my lower back

Answers to Interview Questions

(1) What seems to be bothering you?

I've been having this awful stomach pain. And um it's actually starting to worry me

(2) How long has this been happening?

About two months

(3) Do you feel this pain anywhere else?

Yes, in my back

(4) Can you tell me how long the pain lasts?

It's a constant, dull aching pain all day

(5) Where is the pain located?

In the small of my back

(6) Is there anything that makes the pain better or worse?

The asprin gives me a little relief from the back pain, but it still bothers me. Eating also helps

(7) On a scale of 1-10, with 10 being the highest, how would you rate the pain?

8

(8) Have you ever had any blood transfusions? When?

No

(9) Have you had any past major illnesses?

No

(10) Do you have any significant family history of serious illness?

My mother died of a heart attack when I was 20 and my father died of colon cancer

(11) Do you have any idea of what might be causing the pain?

No, I am not sure

(12) Have you ever had any surgeries?

No

(13) Have you ever been hospitalized for anything?

No

(14) Can you tell me what medications you are taking?

1 Zestril per a day for high blood pressure and (prescription), Asprin for back spasms (over the counter), and tums for my stomach ache

(15) Do you smoke? Yes

Protocol Number: NAWCTSD.2011.0008, TSD 192

Protocol Title: Virtual Humans vs. Prerecorded Humans in Standardized Patient Performance Assessment

(16) How often do you drink?

2-3 bottles per week

(17) Is there anything that you are allergic to?

Allergic to penicillin-causes a rash, shellfish – causes swollen lips

(18) Are your immunizations up to date?

Yes

(19) How is your diet? Do you eat well?

I usually eat a quick breakfast, occasionally have fast food for lunch, and usually have a healthy dinner at home. Overall, I eat well.

Protocol Number: NAWCTSD.2011.0008, TSD 192

Protocol Title: Virtual Humans vs. Prerecorded Humans in Standardized Patient Performance Assessment

Perceived Ease of Use and Open Ended Feedback Survey

For the following please fill a number on a scale of 1-5 as appropriate in the blanks for each question. Leave the answer blank if you don't understand the question. The survey questions are taken from Davis' perceived usefulness, perceived ease of use, and user acceptance of Information Technology.

Strongly Disagree Somewhat Disagree Disagree Neutral Agree Somewhat Agree Strongly Agree
1-----2-----3-----4-----5-----6-----7

Perceived Ease of Use

- ☐ I often get confused when I use interactive computer based training
- ☐ I make errors often when using interactive computer based training
- ☐ Interactive computer based training is often frustrating
- ☐ I need additional instructions to use interactive computer based training
- ☐ Using interactive computer based training requires a lot of mental effort
- ☐ I find it easy to recover from errors in interactive computer based training
- ☐ Using interactive computer based training is rigid and inflexible to interact with
- ☐ I find it easy to get interactive computer based training to do what I want it to do
- ☐ The interactive computer based training often acts in unexpected ways
- ☐ I find it cumbersome to use the interactive computer based training
- ☐ It is easy to remember how to perform tasks using interactive computer based training
- ☐ The interactive computer based training provides helpful guidance in performing tasks
- ☐ Overall, I find interactive computer based training easy to use

What would you add to make this better? _____

What did you like about this Training experience? _____

What did you dislike about this Training experience? _____

Additional feedback about your experience?

Protocol Number: NAWCTSD.2011.0008, TSD 192

Protocol Title: Virtual Humans vs. Prerecorded Humans in Standardized Patient Performance Assessment

Post Survey

POSITIVE AND NEGATIVE AFFECT SCHEDULE (PANAS)

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now, that is, at the present moment. Use the following scale to record your answers.

1-----	2-----	3-----	4-----	5-----	6-----	7-----
Not at all	close to nothing	a little	moderately	moderately more	a lot more	extremely
_____	interested				_____	irritable
_____	distressed				_____	alert
_____	excited				_____	ashamed
_____	upset				_____	inspired
_____	strong				_____	nervous
_____	guilty				_____	determined
_____	scared				_____	attentive
_____	hostile				_____	jittery
_____	enthusiastic				_____	active
_____	proud				_____	afraid

For the following please fill a number on a scale of 1-7 as appropriate in the blanks for each question. Leave the answer blank if you don't understand the question. The survey questions are taken from Davis' perceived usefulness, perceived ease of use, and user acceptance of Information Technology.

Strongly Disagree	Somewhat Disagree	Disagree	Neutral	Agree	Somewhat Agree	Strongly Agreed
1-----	2-----	3-----	4-----	5-----	6-----	7-----

Perceived Usefulness

- __ Training would be difficult to complete without interactive computer based training
- __ Using interactive computer based training gives me more control of my training
- __ Using interactive computer based training improves my performance
- __ Using interactive computer based training addresses my training needs
- __ Using interactive computer based training saves me time
- __ Using interactive computer based training allows me to complete my tasks more quickly
- __ Using interactive computer based training enhances my effectiveness
- __ Using interactive computer based training makes it easier to learn
- __ Overall, I find interactive computer based training useful in learning

Protocol Number: NAWCTSD.2011.0008, TSD 192

Protocol Title: Virtual Humans vs. Prerecorded Humans in Standardized Patient Performance Assessment

Post Participation Information

Interactive Standardized Patient Training

We realize you have several opportunities to aid in research, and your assistance and enthusiasm in my research has been a large asset to my results and discoveries. The purpose of this research is to investigate the effectiveness of Virtual Humans vs Pre-Recorded Humans in a Standardized Patient Performance Assessment. The data that you provided will assist in developing better technologies that will improve the quality and cost effectiveness of Standardized Patient and Clinical Skills education .

Talking interactively with a computer or being recorded can occasionally be uncomfortable. If involvement in a particular part or overall part of the study gave you this sensation, feel free to let us know before leaving the lab. Your satisfaction with the study and well being are of our utmost concern.

If you could please maintain confidentiality regarding the purpose of this study, what you witnessed, discovered, and learned in this encounter, it would be greatly be appreciated. This discretion will assist in preserving the integrity of this research study.

If you have any questions, comments, or issues that you feel that you would like to discuss, please contact Joel Palathinkal. His email is joelpalathinkal@gmail.com. We would also like for you to provide your feedback to aid in future research participation. This form is attached and all directions needed are on the form.

Thank you very much for participating!

Joel Palathinkal

UCF PhD Candidate

Institute for Simulation and Training

APPENDIX C: SLIDES FOR INTRODUCTION AND PRACTICE SESSION

Press ENTER to Begin

Keep pressing **ENTER** at the end of each slide to continue throughout the module

Press **ENTER** twice if the slides don't continue immediately

Press ENTER to Continue

Technical Guidance

Audio:
-If you didn't hear what was said the first time,
-Double click on the speaker icon
usually located on the lower left side
-You can Test it by double clicking the icon (red arrow points)

Slides:
-You can normally continue to the next slide
by pressing **ENTER**
-Press Enter once more if it still doesn't continue

Further Technical Questions:
-For any other technical questions
-Please notify your research coordinator
-They will work to fix it immediately


9

Standardized Patient Training

Welcome to Standardized Patient Training. This training session will allow you to learn the roles of a Standardized Patient and develop a better knowledge of a Standardized Patient Case and Interview

Press ENTER to Continue

What is a Standardized Patient (SP)?



An SP is an Actor:
-Representative of a patient
-Possessive knowledge of an SP case
-Able to portray medical scenarios
-Searches for clues in Standardized Patient Case packet

6

Standardized Patient (SP) Roles

- Portraying a patient in a medical case
- Possessing medical knowledge of an SP case
- Evaluating the medical student performance
- SPs are very important in training medical students
 - role play training
 - clinical exams
- The better the SP, the better the medical student

9

Training Objectives of this Module

Training users to learn the role of an SP

Training users to learn an SP case

Participation in an interactive SP interview

- interactive training
- correctly answering questions related to the case

9

Complete Study Forms

- At this time, please Refer to the paper packet that was given to you
- Please read, sign, and complete pages 1-5
- Notify the study coordinator once this is done

Press ENTER to Continue

Standardized Patient Case

- The next slides will be training on the Standardized Patient Case
- Please read the narrated case that describes the scenario on the following slides
- Press ENTER to start the Standardized Patient Case Review

Press ENTER to Continue

Scenario

This is the scenario you must know to perform well in the interview you will participate in shortly. You are being seen in clinic because you have been having pain in the middle of your stomach for the past two months. You have also been having low back pains for two or three months.

You have been taking a Zestril once a day for your high blood pressure for the past 4 years. Your blood pressure has been well controlled on the medication. You take the medicine consistently. You do not have any symptoms associated with your high blood pressure. In fact, it was discovered on an annual physical. High blood pressure runs in your family. You come to the clinic twice a year for your prescription and blood pressure checks. You also check your blood pressure at Publix and Eckerd's from time to time and it is always in the normal range.

Press ENTER to Continue

Scenario

Your back pain started two to three months ago (you're not sure) after you injured it at work while lifting boxes. You work in the mail room of the post office. The back pain is located in the small of your back. The pain is a constant, dull aching located deep in the center of your back. You have been able to work, but the pain gets worse as the day progresses so that by mid afternoon you can't sit comfortably. You have curtailed your activities and taken aspirin. You take adult strength aspirin, two pills, two to three times a day. The aspirin gives you a little relief from the back pain, but the back pain still bothers you. Sitting for prolonged periods and bending over make the pain worse. Lying down is the most comfortable position for relief. You tend to shift your weight to the right side when sitting and standing. On a scale of 1 to 10, with 10 being the highest, you would rate your pain as 8.

You have been taking aspirin for as long as your back has been hurting. Your stomach has been hurting for about the same amount of time, maybe a little less. When you take the stomach pain feels better. [Press ENTER to Continue](#)

Scenario

You have generally been in good health. You have had no surgeries, blood transfusions, serious illnesses, or injuries.

Hospitalizations: No hospitalizations.

Allergies: Penicillin - causes a rash.
Shellfish - swollen lips.

Chronic illnesses: High blood pressure. Diagnosed 4 years ago.

Immunizations: up to date. Medications: Zestril 1 per day.

Over the Counter Meds: Aspirin (non-steroid) for back spasms, turns for stomach ache.

Illicit drugs: none.

Social History:
You are married and in a monogamous relationship. [Press ENTER to Continue](#)

Scenario

Education: BS in Business
Employment: Post Office

Exercise: You walk for about 30 minutes a day.

Tobacco: You smoke a half pack of cigarettes a day, and you started smoking when you were 16.

Alcohol: 2-3 bottles of beer a week.

Diet: You eat a quick breakfast, occasional fast food for lunch, and usually have dinner at home. Overall, you eat well.

Sexual History:
You have been married for 10 years. You are sexually monogamous. You have never had a sexually transmitted disease. [Press ENTER to Continue](#)

Scenario

Family History:
Both your parents are gone. Your Dad died a couple years ago, and your Mom died when you were twenty.
Mother had a heart attack.
Father died of colon cancer.


Presentation: Emotional Tone
You are concerned about the back problem and stomach problems because they are not resolving on their own. You are particularly concerned because your father died of colon cancer which presented first as stomach problems.
You will answer questions cooperatively regardless of the interview style. You are neatly groomed, dressed in pants and a nice shirt. [Press ENTER to Continue](#)

Thank You

- Thank you for completing the Standardized Patient Case Review
- Please notify the study coordinator that you have completed this section.

[Press ENTER to Continue](#)

Practice Session



Become Familiar with:

- SP Case
- Navigating through slide
- Recording your voice
- Experience on Computer based training for SP interviews

[Press ENTER to Continue](#)

Lets Start the Practice Session!!

Now you are ready to begin the SP interview Practice Session. You can look at the answers or the patient that was given to you if you get stuck on any interview questions.

Character

Opening Statement
I've been having this awful stomach pain, and it's starting to worry me.

History

1. I take a Zestril once a day to control my blood pressure.
2. My blood pressure was diagnosed 4 years ago.
3. I have been having awful stomach pain.
4. The stomach pain is gnawing, and sometimes it's sharp, especially at night.
5. I do not have any symptoms related to my blood pressure (headache, chest pain).
6. My back pain started two or three months ago.
7. The pain is constant, deep, and aching.
8. The pain is in my lower back.

[Press ENTER to Continue](#)

Reminder!!

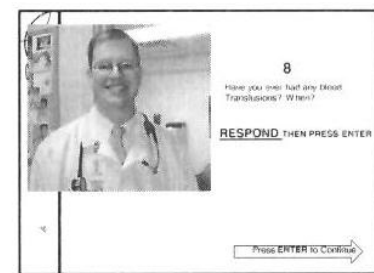
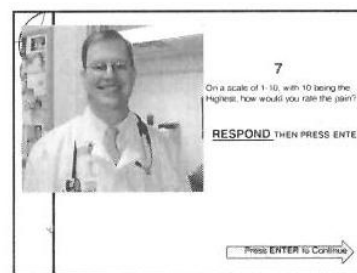
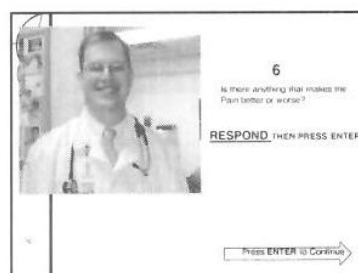
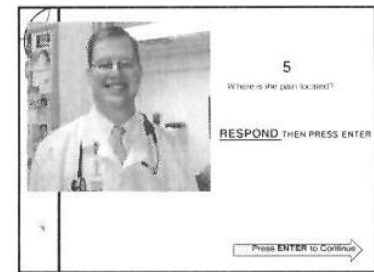
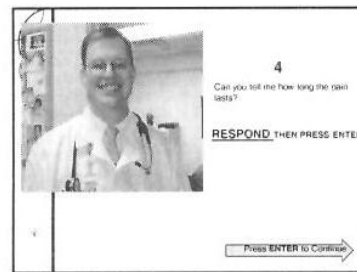
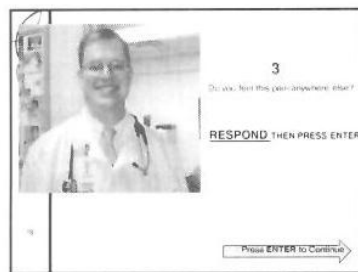
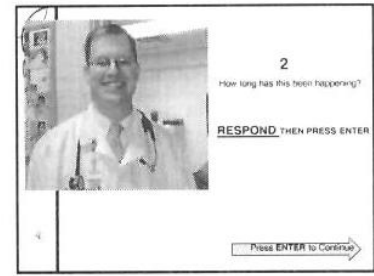
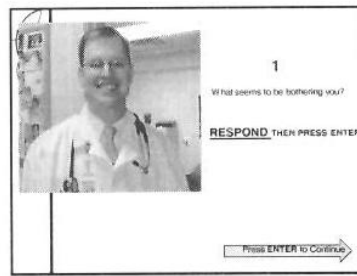
- Pages 6-9 of your study packet contain the case that you just read along with the answers to questions that you will be asked shortly.
- An interactive Medical student is about to ask you some interview questions regarding the case that you read.
- Answer them while pretending that you are a patient.
- However, try to answer these questions to the best of your ability without looking at the answers.

[Press ENTER to Continue](#)

Lets Meet Tony the Medical Student!!

- Answer his questions as if you were having a conversation with him with your voice.
- The next slides will require you to speak into the microphone and respond.
- After you respond, press enter to move to the next question.
- Keep going through the interview until you have reached the end.

[Press ENTER to Continue](#)



9

Have you had any (and major) illnesses?

RESPOND THEN PRESS ENTER

Press ENTER to Continue

10

Do you have any significant family history of serious illness?

RESPOND THEN PRESS ENTER

Press ENTER to Continue

11

Do you have any idea what might be causing the pain?

RESPOND THEN PRESS ENTER

Press ENTER to Continue

12

Have you ever had any surgeries?

RESPOND THEN PRESS ENTER

Press ENTER to Continue

13

Have you ever been hospitalized for anything?

RESPOND THEN PRESS ENTER

Press ENTER to Continue

14

Can you tell me what medications you are taking?

RESPOND THEN PRESS ENTER

Press ENTER to Continue

15

Do you smoke?

RESPOND THEN PRESS ENTER

Press ENTER to Continue

16

How often do you drink?

RESPOND THEN PRESS ENTER

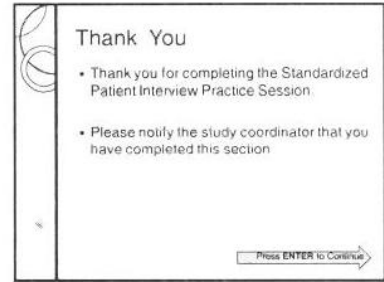
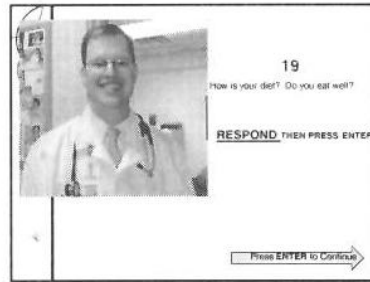
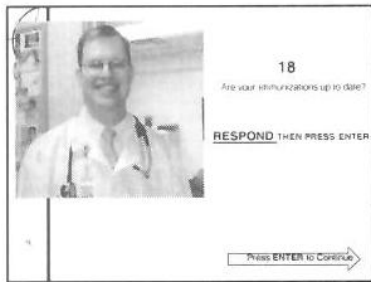
Press ENTER to Continue

17

Is there anything that you are allergic to?

RESPOND THEN PRESS ENTER

Press ENTER to Continue



APPENDIX D: SLIDES FOR VIRTUAL HUMAN PERFORMANCE ASSESSMENT

Performance Assessment

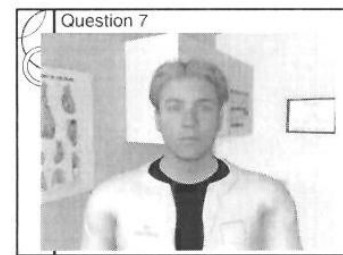
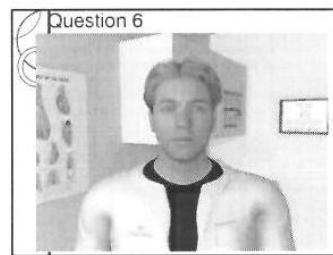
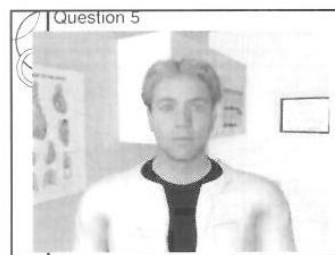
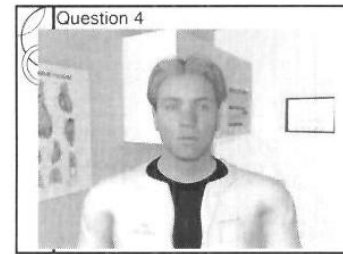
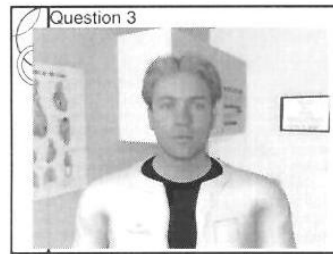
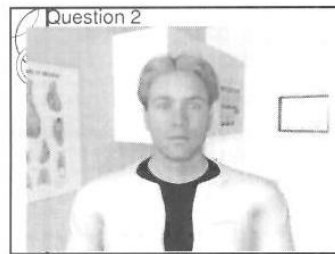
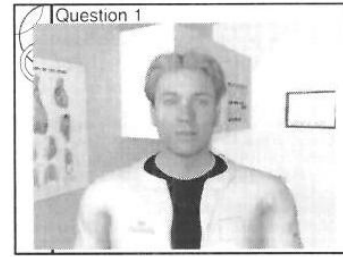
- You will now be given a simulated interview with a medical student.
- You will be recorded during these sections.
- The next slides will be a series of questions.
- This time, you cannot refer to the answers.
- Answer them to the best of your ability, by speaking your responses based on what you learned in training and in practice.
- If a virtual medical student freezes up on the screen, double click on it to refresh it.
- Notify the research coordinator if there are technical difficulties.

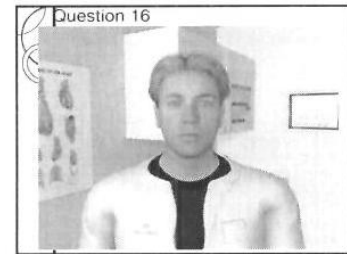
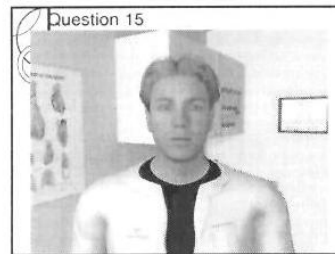
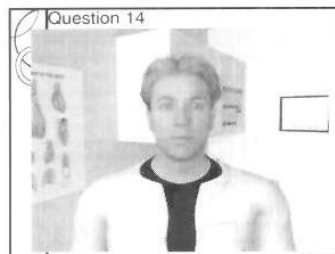
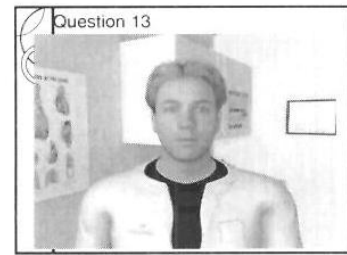
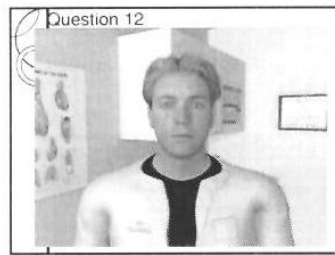
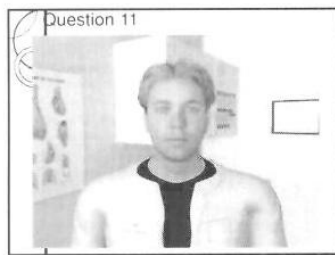
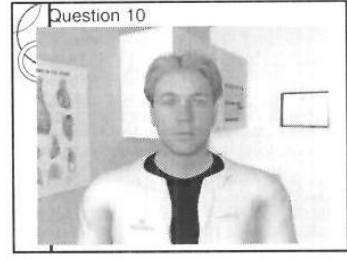
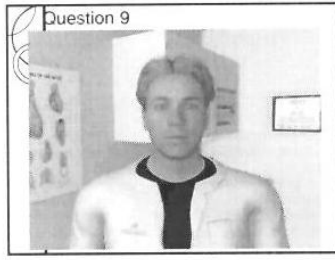
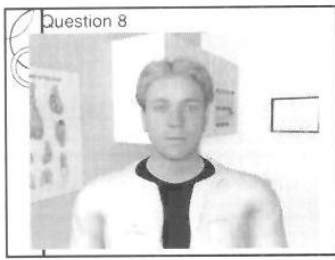
Press ENTER to Continue

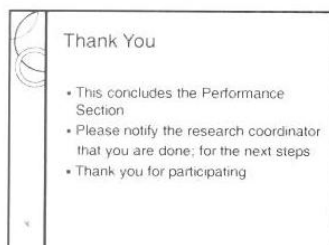
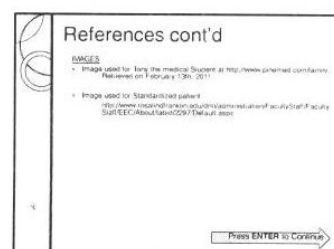
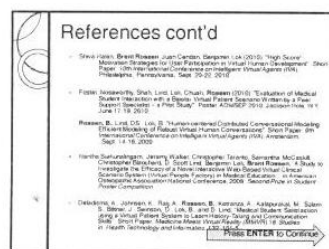
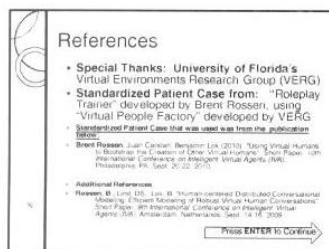
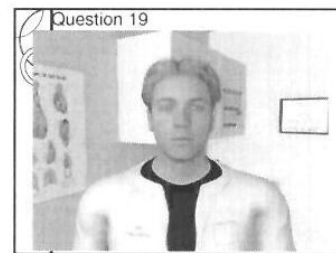
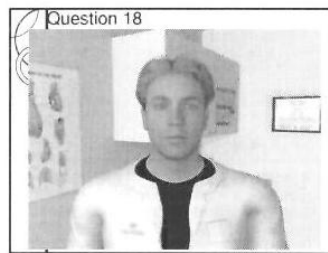
Performance Assessment

- Please Press ENTER to begin the Standardized Patient Interview.

Press ENTER to Continue







APPENDIX E: SLIDES FOR PRE-RECORDED HUMAN PERFORMANCE
ASSESSMENT

Performance Assessment

- You will now be given a simulated interview with a medical student
- You will be recorded during these sections
- The next slides will be a series of questions
- This time, you cannot refer to the answers
- Answer them to the best of your ability, by speaking your responses based on what you learned in training and in practice.
- If a virtual medical student freezes up on the screen, double click on it to refresh it
- Notify the research coordinator if there are technical difficulties


Press ENTER to Continue

Performance Assessment


- Please Press ENTER to begin the Standardized Patient Interview

Press ENTER to Continue


Question 1




Question 2



Question 3




Question 4




Question 5

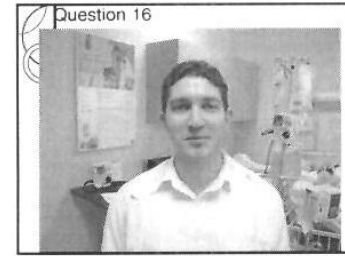
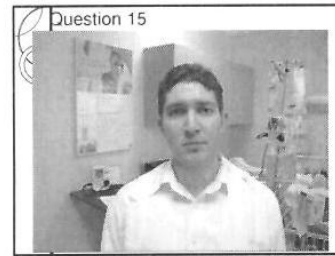
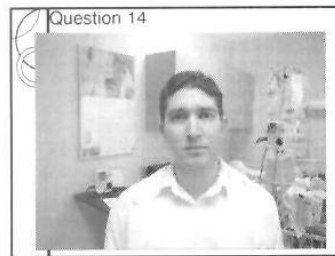
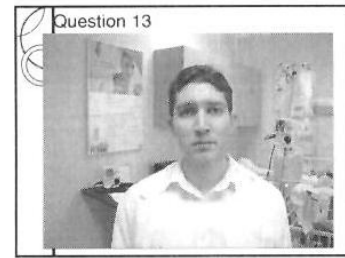
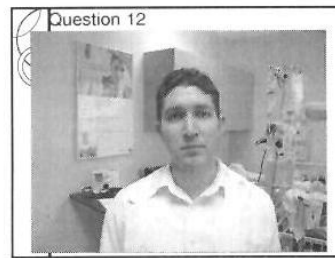
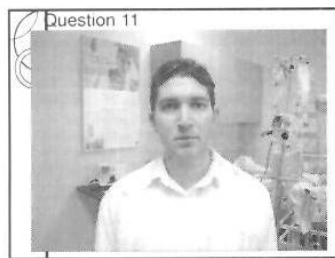
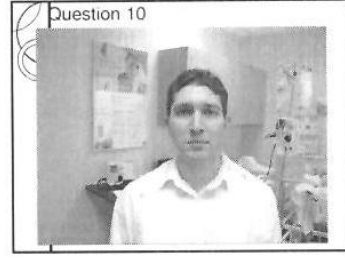
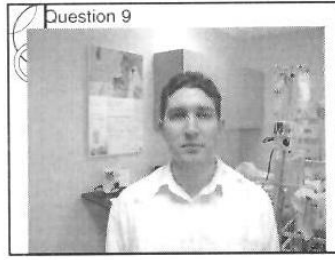
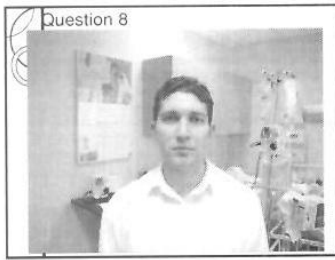


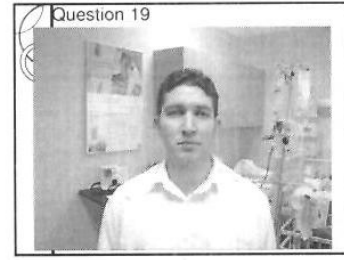
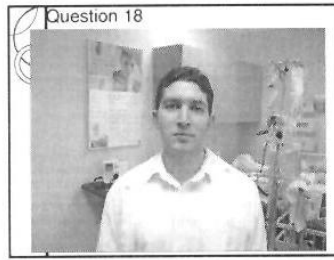
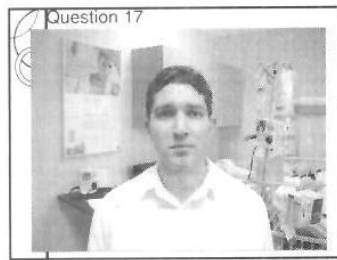
Question 6



Question 7







References

- **Special Thanks:** University of Florida's Virtual Environments Research Group (VERG)
- **Standardized Patient Case from:** "Roleplay Trainer" developed by Brent Rossen, using "Virtual People Factory" developed by VERG
- **Standardized Patient Case that was used was from the publication below:**
- **Brent Rossen, Juan C. Garcia, Benjamin Lee (2010),** "Moving Virtual Humans to Boosting the Impact of Other Virtual Humans," *Virtual Paper*, ACM International Conference on Intelligent Virtual Agents (IVA), Philadelphia, PA, Sept. 20-22, 2010.
- **Additional References:**
- **Rossen, B., Lee, D.S., Lee, B.,** "Human-centered Distributed Conversational Modeling: Efficient Modeling of Human-Virtual Human Conversations," *Short Paper*, ACM International Conference on Intelligent Virtual Agents (IVA), Amsterdam, Netherlands, Sept. 19-21, 2009.

Press ENTER to Continue

References cont'd

- **Shen, Haido, Brent Rossen, Juan C. Garcia, Benjamin Lee (2010),** "Tight-Score! Research Strategies to Use Participation in Virtual Human Development," *Short Paper*, ACM International Conference on Intelligent Virtual Agents (IVA), Philadelphia, Pennsylvania, Sept. 20-22, 2010.
- **Foster, Rosemary, Shen, Haido, Lee, B., Crain, Rossen (2010),** "Evaluation of Medical Student Interaction with a Bipolar Virtual Patient Scenario in a Role-Play Patient Scenario" - *Final Study*, *Poster*, *CHI 2010*, Jacksonville, FL, 2-6, 2010.
- **Rossen, B., Lee, D.S., Lee, B.,** "Human-centered Distributed Conversational Modeling: Efficient Modeling of Human-Virtual Human Conversations," *Short Paper*, ACM International Conference on Intelligent Virtual Agents (IVA), Amsterdam, Sept. 19-21, 2009.
- **Nardi, Barbara, J., Henry, William, Christopher, Dorian, Samantha, McCall, Christopher, Schwartz, J., Scott, Lisa, Benjamin, Lee, Brent, Rossen, B.,** "A Study to Investigate the Effect of a Role-Playing in a Distributed Virtual Clinical Simulation System," *Virtual People Factory in Medical Simulation*, *Online*, *Association for the Advancement of Artificial Intelligence (AAAI) Symposium on Abstraction, Reformulation and Approximation (SARA)*, 2008, Boston, MA, 2008.
- **Choudhury, A., Johnson, A., Ray, A., Rossen, B., Haido, B., Haido, B., Lee, D.S., Lee, B.,** "Human-centered Distributed Conversational Modeling: Efficient Modeling of Human-Virtual Human Conversations," *Short Paper*, ACM International Conference on Intelligent Virtual Agents (IVA), Amsterdam, Netherlands, Sept. 19-21, 2009.

Press ENTER to Continue

References cont'd

- **Image used for: Tanya the medical Student at** <http://www.patient.com/family> Retrieved on February 13th, 2011
- **Image used for: Standardized patient** <http://www.standardspatient.com/administrative/facultyStaff/facultyStaff.cfm?showImage=2251&ext=jpg>

Press ENTER to Continue

Thank You

- This concludes the Performance Section
- Please notify the research coordinator that you are done, for the next steps
- Thank you for participating

APPENDIX F: ADDITIONAL DATA/ARTIFACTS

The demographic data is utilized to understand if there are any significant differences among the demographic parameters. The Multiple ANOVA is conducted to analyze demographics.

Analysis of Variance for Assessment score, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Training method	1	1.923	1.057	1.057	0.11	0.745
Gender	1	40.297	11.675	11.675	1.18	0.283
Age	2	21.717	33.579	16.789	1.70	0.195
Race	4	129.068	116.032	29.008	2.94	0.032
Advanced degree (Yes/No)	1	1.063	3.863	3.863	0.39	0.535
Acting Experience (Yes/No)	1	1.442	1.317	1.317	0.13	0.717
Healthcare experience (Yes/No)	1	23.002	25.805	25.805	2.62	0.114
Bilingual (Yes/No)	1	6.735	6.735	6.735	0.68	0.414
Error	39	384.523	384.523	9.860		
Total	51	609.769				

The multiple ANOVA suggests that only “Race” is the significant factor indicating that at least one Race is different than others. In our study, there are five different Race types as Caucasian, Asian, Hispanic, African American, and Native American. But we do not know from ANOVA which Race different than other Races. To conduct this analysis, we conducted Tukey’s test. The output is shown below.

Tukey 95% Simultaneous Confidence Intervals. All Pairwise Comparisons among Levels of Race. Individual confidence level = 99.33%

an subtracted from:				Race = African American subtracted from:			
	Lower	Center	Upper	Race	Lower	Center	Upper
erican	-2.580	2.250	7.080	Caucasian	-1.673	1.942	5.558
	0.142	4.192	8.243	Hispanic	-1.406	2.750	6.906
	0.461	5.000	9.539	Native American	-3.736	5.750	15.236
frican	-1.660	8.000	17.660				
-----+-----+-----+-----+-----				Race	-----+-----+-----+-----		
erican	(----*----)			Caucasian	(---*---)		
	(---*---)			Hispanic	(---*---)		
	(----*----)			Native American	(-----*-----		
frican	(-----*-----))	-----+-----+-----+-----		
-----+-----+-----+-----+-----				-----+-----			
	-10	0	10	20	-10	0	10

Table 8: Additional Tukey Analysis

Charts of Pre and Post survey (For PANAS scale)

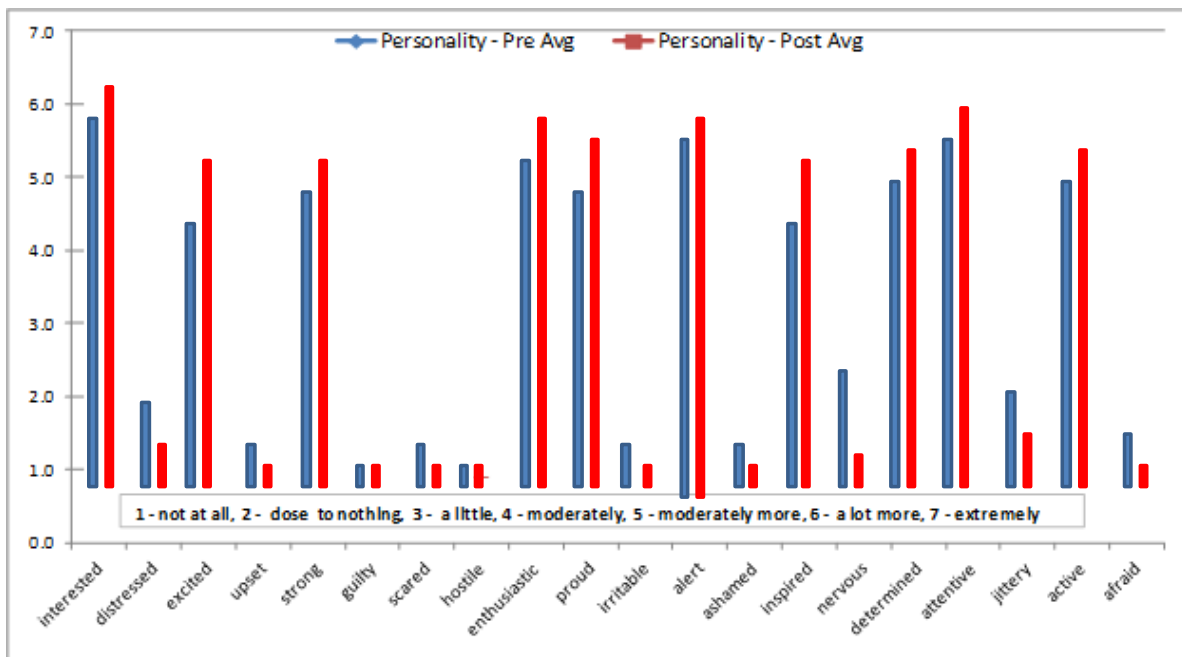


Table 9: Pre and Post Survey PANAS scales

The chart in shows the subject's feeling at the present time using a PANAS scale before and after the study participation. The characteristics are on horizontal axis (X axis) and score (between 1 to 7) is on vertical axis. The data showed positive movement for each category in post survey. For example, subjects are more interested after the study, and subjects are less distressed after the study.

Chart of Perceived Usefulness Survey

The chart for perceived usefulness is plotted below. The description is assigned acronyms as shown below.

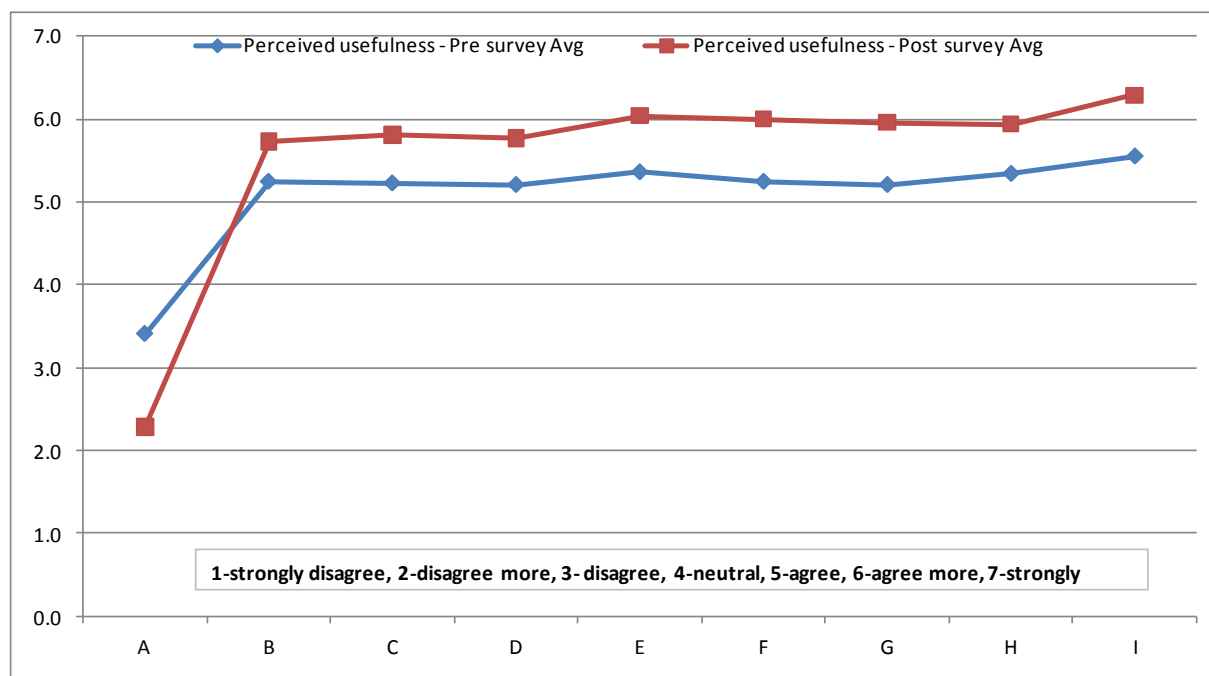


Table 10: Perceived Usefulness

DESCRIPTION	KEY
Training would be difficult to complete without computer based performance assessments	A
Using computer based performance assessments gives me more control of my training	B
Using computer based performance assessments improves my performance	C
Using computer based performance assessments addresses my training needs	D
Using computer based performance assessments saves me time	E
Using computer based performance assessments allows me to complete my tasks more quickly	F
Using computer based performance assessments enhances my effectiveness	G
Using computer based performance assessments makes it easier to learn	H
Overall, I find computer based performance assessments to be useful in learning	I

The data in the figure shows an overall positive attitude towards computer based performance assessments.

Chart for Perceived ease of use

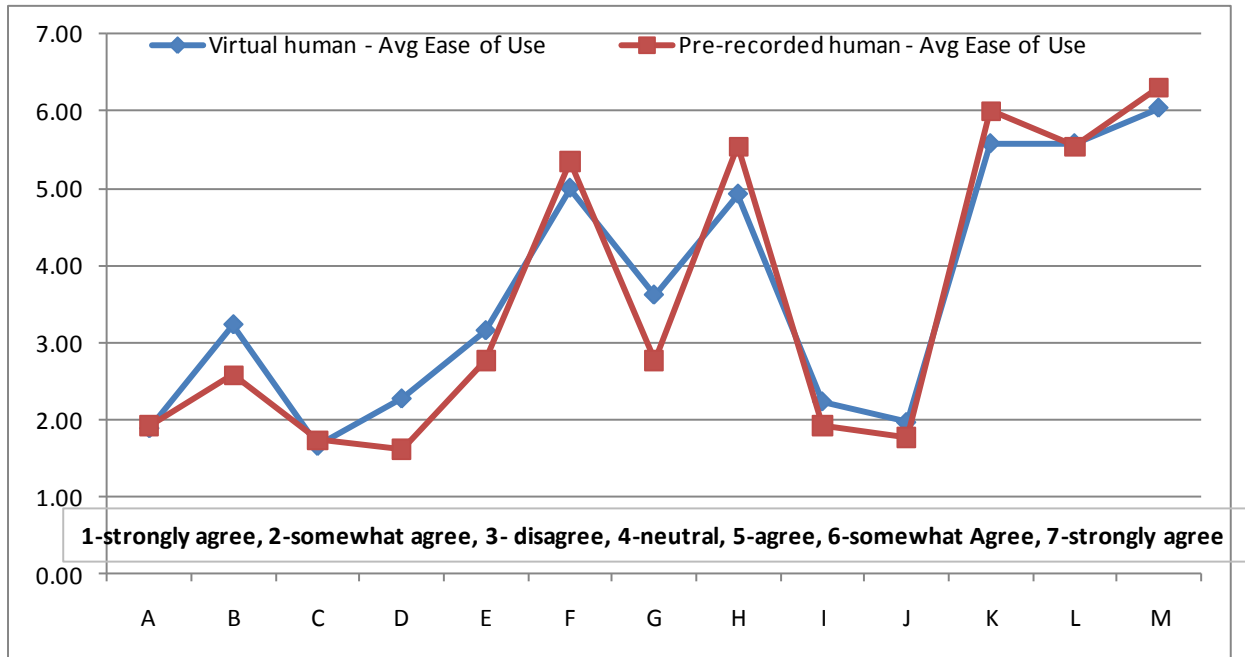


Table 11: Perceived Ease of Use

DESCRIPTION	KEY
I often get confused when I use computer based performance assessments	A
I make errors often when using computer based performance assessments	B
Computer based performance assessments are often frustrating	C
I need additional instructions to use computer based performance assessments	D
Using computer based performance assessments require a lot of mental effort	E
I find it easy to recover from errors in computer based performance assessments	F
Using computer based performance assessments are rigid and inflexible to interact with	G
I find it easy to get computer based performance assessments to do what I want it to do	H
The computer based performance assessments often acts in unexpected ways	I
I find it cumbersome to use computer based performance assessments	J
It is easy to remember how to perform tasks using computer based performance assessments	K
Computer based performance assessments can provide helpful guidance in performing tasks	L
Overall, I find computer based performance assessments easy to use	M

Standardized Patient Questionnaire

- 1) Which medical student below would you prefer to train with for a Standardized Patient case? Why are you choosing that option?

*either one,
but I think
a real person
is easier to
respond*



#1 (Simulated) Virtual Human #2 (Real-Life) Pre-Recorded Human

- 2) From above, would a Simulated Virtual Human (#1) or Pre-Recorded Human (#2) capture your attention more? Why?
- 3) Do you think using a simulation/computer based session saves you time in learning a standardized patient case? Why?
- 4) What current medical simulation technologies for Standardized Patient Training have you used/been aware of?
- 5) What are your biggest obstacles in learning to be a Standardized Patient? How do you overcome them?
- 6) What could be done to improve Standardized Patient Education?
- 7) How did you learn to become a Standardized Patient? *Debbie Whiston told us about it*
- 8) What would help you learn your lines better, portray a patient more realistically, and improve your memory?
- 9) Are you an Actor? If so, for how long and to what level of proficiency (amateur, experienced, professionally trained)?
I guess we are now
- 10) What could medical students/Standardized Patient educators do to help in the process?

Figure 39: Standardized Patient Questionnaire 1

Standardized Patient Questionnaire

- 1) Which medical student below would you prefer to train with for a Standardized Patient case? Why are you choosing that option?



#1 (Simulated) Virtual Human #2 (Real-Life) Pre-Recorded Human

If find a real person easier to pay attention to

- 2) From above, would a Simulated Virtual Human (#1) or Pre-Recorded Human (#2) capture your attention more? Why?

#2 Because the pre-recorded human is more "believable" to follow in language, expression and emotional content

- 3) Do you think using a simulation/computer based session saves you time in learning a standardized patient case? Why?

I think if you are a person who learns better and retains more information by visualization, then a simulation/computer based session would be advantageous to utilize

- 4) What current medical simulation technologies for Standardized Patient Training have you used/been aware of?

Pre-recorded cases streamed through the computer

- 5) What are your biggest obstacles in learning to be a Standardized Patient? How do you overcome them?

Memorizing "minor" numerical details (dates, units of measure

- 6) What could be done to improve Standardized Patient Education?

for prescriptions, ages) closer times between some training modules and actual sessions

- 7) How did you learn to become a Standardized Patient?

I saw a news story on Channel 9 about the need for volunteers

- 8) What would help you learn your lines better, portray a patient more realistically, and improve your memory?

more practice sessions with a preceptor or monitor to give feedback on areas of my performance that may need improvement

- 9) Are you an Actor? If so, for how long and to what level of proficiency (amateur, experienced, professionally trained)?

No, but I enjoy acting like an actor (amateur)

- 10) What could medical students/Standardized Patient educators do to help in the process?

*Medical students need to stay in character at all times
SPs need to make sure they are giving consistent and correct portrayals of their assigned characters*

Figure 40: Standardized Patient Questionnaire 2

REFERENCES

- Adjoudani, A., and C. Benoit. 1996. On the integration of auditory and visual parameters in an HMM-based ASR. In D.G. Stork and M.E. Hennecke, eds., *Speechreading by humans and machines: Models, systems, and applications*, 461-472. Berlin: Springer-Verlang
- Andre, E, and T. Rist. 1995. Generating coherent presentations employing textual and visual material. *Artificial Intelligence Review* 9(2-3):147-165. (Special issue on the Integration of Natural Language and Vision Processing.)
- Azarbeyejani, A., C. Wren, and A. Pentland. 1996. Real time 3-D tracking of the human body. In Proceedings of IMAGE'COM 96 (Bordeaux, France), May
- Badler, N., M. Palmer. And R. Bindiganavale. 1999. Animation control for real time virtual humans. *Communications of the ACM* 42(7):65-73.
- Baker, Ann C, Jensen, Patricia J, & Kolb, David A. (2005). Conversation as Experiential Learning. *Management Learning*, 36(4), 411-427. Retrieved November 4, 2010, from ABI/INFORM Global. (Document ID: 944428841).
- Baker, A. C., Jensen, P. J. and Kolb, D. A. (1997) 1In Conversation: Transforming Experience into Learning'. *Simulation and Gaming*28(1): 6-12
- Ball, G., Ling, D Kurlander, J Miller, D. Pugh, et al 1997. Lifelike computer characters: The persona project at Microsoft Research. In J.M. Bradshaw, ed., *Software agents*, 191-222. Menlo park, Calif: AAAI Press/The MIT Press.
- Bavelas, J., N Chovil L, Coates, and L. Roe 1995. Interactive gestures. *Dicourse processes* 15:469-489
- Beatty, G. W. 1981 Sequential temporal patterns of speech and gaze in dialogue. In T.A Sebeok and J Umiker-Sebeok, eds. *Nonverbal communication, interaction, and gesture: Selections from Semiotica*, 298-320. The Haugue: Mouton
- Behrend, Tara S., Thompson, Lori Foster (2011). Similarity Effects in Online Training: Effects with Computerized Trainer Agents. *Journal of Computers in Human Behavior*. Volume 27, Issue 3. Elsevier Science Publishers B.V. Amsterdam, Netherlands.
- Bolan, CM. 2003 Incorporating the *experiential learning* theory into the instructional design of online courses. *Nurse Educator* 2003 Jan-Feb; Vol. 28 (1), pp. 10-4.
- Bolt, R A 1980. Put-That-There: Voice and Gesture at the graphics interface. *Computer Graphics* 14 (3):262-270.
- Bradley, C., Webb, T., Schmitz, C., Chipman, J., & Brasel, K. (2010). Structured teaching versus experiential learning of palliative care for surgical residents. *American Journal Of Surgery*, 200(4), 542-547. Retrieved from MEDLINE database.
- Chronbach's Alpha, *Wikipedia* retrieved on March 4, 2010 from

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Churchill, E. F., L. Cook, P. Hodgson, S. Prevost, and J. W. Sullivan. 1999. Designing embodied conversational agent allies. A scenario based approach. FX Palo Alto Technical Report FXPAL-TR-99-023
- Conkey, A Curtis (2009). Video Based Soft Skills Training dissertation for Doctor of Philosophy in Modeling and Simulation in the College of Sciences at the University of Central Florida Orlando, Florida
- Cassell, Justine. (2000). Embodied Conversational Agents. MIT Press
- Corty, Eric. Using and Interpreting Statistics: A practical text for the health, behavioral, and social sciences. Mosby; 1 Edition (September 25, 2006).
- Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* 13 (3), 319-340 (1989)
- Dickerson, R., et al.: Evaluating a Script-Based Approach for Simulating Patient-Doctor Interaction. In: SCS 2005 International Conference on Human-Computer Interface Advances for Modeling and Simulation, pp. 79–84 (2005)
- Diener, E., Larsen, R. J., Levine, S., & Emmons, R. A. (1985). Intensity and frequency: Dimensions underlying positive and negative affect. *Journal of Personality and Social Psychology*, 48, 1253-1265.
- Duncan, S. 1974. Some signals and rules for taking speaking turns in conversations. In S. Weitz, ed., *Nonverbal communication*. New York: Oxford University Press
- Errichetti A, Boulet JR. 2006 Oct;81(10 Suppl):S91-4. Comparing traditional and computer-based training methods for standardized patients. *Academic Medicine: Journal of the Association of American Medical Colleges*.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3. A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Feldman, M. (2008). *Controlling Our Emotion at Work: Implications for Interpersonal and Cognitive Task Performance in a Customer Service Simulation*. Unpublished Dissertation, University of Central Florida, Orlando, Florida.
- First Aid for the usmle step 2cs
- Foster, Noseworthy, Shah, Lind, Lok, Chuah, **Rossen** (2010). "Evaluation of Medical Student Interaction with a Bipolar Virtual Patient Scenario Written by a Peer Support

- Specialist – a Pilot Study". Poster. ADMSEP 2010, Jackson Hole, WY, June 17-19, 2010.
- Funke, L., & Booth, J.E. (1961). *Actors Talk About Acting*. Random House: New York
- Furman GE. 2008 Dec;24(12):651-5. The role of standardized patient and trainer training in quality assurance for a high-stakes clinical skills examination. *Kaohsiung Journal of Medical Science*.
- Fussell, Holly E. , Lewy, Colleen S. and McFarland, Bentson H.(2009) 'Evaluating and Training Substance Abuse Counselors: A Pilot Study Assessing Standardized Patients as Authentic Clients', *Substance Abuse*, 30: 1, 47 — 60
- Halan, Shiva, Rossen, Brent, Cendan, Juan, Lok, Benjamin (2010). "High Score! – Motivation Strategies for User Participation in Virtual Human Development". Short Paper. *10th International Conference on Intelligent Virtual Agents (IVA)*, Philadelphia, Pennsylvania, Sept. 20-22, 2010.
- Hall, Judith A. 2009. Jan 6th. Patient Education and Counseling. Observer Rated Rapport in interactions between medical students and standardized patients.
- Heeter, C. (1992). Being there: the subjective experience of presence. *PRESENCE: Teleoperators and Virtual Environments*, 1(2), 262-271.
- Hooks, Ed., *Acting for Animators: A complete guide to animators: A complete guide to performance animation*, revised edition (Portsmouth, NH: Heinemann, 2003)
- Ieronutti, L., & Chittaro, L. (2007). Employing virtual humans for education and training in X3D/VRML worlds. *Computers & Education*, 49(1), 93-109. doi:10.1016/j.compedu.2005.06.007.
- Johnson, W. L., J. Rickel, R Stiles, and A. Munro 1998. Integrating pedagogical agents into virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 7 (6): 523-546.
- Karlowicz, K. (2009). Evaluation of the Urinary Incontinence Scales to Measure Change After Experiential Learning: A Pilot Study. *Urologic Nursing*, 29(1), 40-46. Retrieved from Health Source: Nursing/Academic Edition database.
- Kobak, K., Engelhardt, N., & Lipsitz, J. (2006). Enriched rater training using Internet based technologies: A comparison to traditional rater training in a multi-site depression trial. *Journal of Psychiatric Research*, 40(3), 192-199. doi:10.1016/j.jpsychires.2005.07.012.
- Kolb, D. A. (1984). *Experiential learning: experience as the source of learning and development*: Prentice-Hall Englewood Cliffs, New Jersey.
- Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2001). *Experiential Learning Theory: Previous Research and New Directions. Perspectives on Thinking, Learning, and Cognitive Styles*.

- Kolb D. University of Puget Sound Web site. Available at:
<http://www.ups.edu/aca/advman/mbdevl2.htm>. Accessed March 24, 2003.
- Kolb DA. Experiential Learning. Englewood Cliffs, NJ: Prentice-Hall; 1984;77–78,141.
- Konkle-Parker, D., Cramer, C., & Hamill, C. (2002). Standardized patient training: a modality for teaching interviewing skills. *Journal Of Continuing Education In Nursing*, 33(5), 225-230. Retrieved from MEDLINE database.
- Lagnado, Lucette. These Lagnado Take Dramatic Turns at the Hospital: To Help Train Doctors, Actors Get Gigs Playing Patients; Nabbing a Gout Role. *The Wall Street Journal*. Tuesday August 126th, 2011. Vol CCLVII No. 39
- L. A. Schuh. Education Research: Bias and poor interrater reliability in evaluating the Neurology Clinical Skills Examination. *Neurology* 2009;73;904-908; originally published online Jul 15, 2009; DOI: 10.1212/WNL.0b013e3181b35212
- Lee, J., Triola, M., Gillespie, C., Hanley, K., Zabar, S., Kalet, A., et al. (2008). Working with patients with alcohol problems: A controlled trial of the impact of a rich media web module on medical student performance. *Journal of General Internal Medicine*, 23(7), 1006-1009. doi:10.1007/s11606-008-0557-5
- Lee, E-J., and C. Nass. 1998. Does the ethnicity of a computer agent matter? An experimental comparison of human-computer interaction and computer-mediated communication. In proceedings of the workshop on embedded Conversational Characters Conference (Lake Tahoe, Calif)
- Lester, C James, Deictic and Emotive Communication in Animated Pedagogical Agents (2000). Embodied Conversational Agents. MIT Press
- Lester et al., 1997 Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents. In *Proceedings of CHI '97*.
- Mancini, Flavia, Longo, Mathew. 2011. Visual Distortion of Body Size Modulates Pain Perception. *Psychological Science*
- Massaro, D. W., and M. M Cohen. 1977. Voice onset time and fundamental frequency as cues to the /zi/-/si/dinction. *Perception and Psychphysics* 22:373-382
- May W. 2008 Dec;24(12):640-5. Training standardized patients for a high-stakes Clinical Performance Examination in the California Consortium for the Assessment of Clinical Competence. *Kaohsiung Journal of Medical Science*.
- May, Win. (2008). Training Standardized Patients for a High Stakes Clinical Performance Examination in the California Consortium for the Assessment of Clinical Competence. Standardized Patient Program, Keck School of Medicine of the University of Southern California, USA. Vol 24 • No 12
- Meier, U., R. Stiefelhagen, and J. Yang. 1997. Preprocessing of visual speech under real world conditions. In C. Benoit and R Cambell, eds., *Proceedings of the ESCA*

Workshop on Audio Visual Speech Processing, Cognitive and Computational Approaches (Rhodes, Greece), 113-116.

- Mistry, Pranav. 2009. The Thrilling Potential of Sixth Sense Technology, TED lecture. http://www.ted.com/talks/pranav_mistry_the_thrilling_potential_of_sixthsense_technology.html. Retrieved on August 20, 2011.
- Mott VW. The development of professional expertise in the workplace. *New directions for Adult and Continuing Education* 2000;86 23-31
- Nass, C., and L. Mason. 1990. On the study of technology and task: A variable based approach. In J. Fulk and C. Stienfeld, eds., *Organizations and communication technology*, 46-67. Newbury park, Calif: Sage
- NASA. 1993 CLIPS reference manual. Technical Report, Software Technology Branch, Lyndon B. Johnson Space Center
- O'Connell, Vanessa. At GE, Robo-Lawyers. (2011) Oil and Gas Unit Tests Online Resolution to Control Costs. *Wall Street Journal. Business Technology*.
- Optale, G., Urgesi, C., Busato, V., Marin, S., Piron, L., Priftis, K., et al. (2010). Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. *Neurorehabilitation And Neural Repair*, 24(4), 348-357. Retrieved from MEDLINE database.
- Pettit, H. M. Beth. Research Plans and Joint Medical Simulation Technology (JMST) IPT briefing (2009).
- Phillips, R., & Bonsteel, S. (2010). The faculty and information specialist partnership: stimulating student interest and experiential learning. *Nurse Educator*, 35(3), 136-138. Retrieved from MEDLINE database.
- Plantec, Peter., Kurzweil, Ray. *Virtual Humans: A build it yourself kit, complete with software and step-by-step instructions*. 2003
- Poggi, I. 1991. La comunicazione. In R. Asquini and P. Lucisano, eds., *L'Italiano nella scuola elementare*. Aspetti linguisti. Florence: La Nuova Italia.
- Privateer, P. M. (1999). Academic Technology and the Future of Higher Education: Strategic Paths Taken and Not Taken. *The Journal of Higher Education*, 70(1), 60.
- Rickel, J. and W.L. Johnson 1999a. Animated agents for procedural training in virtual reality perceptionents for procedural training in virtual reality: perception, cognition, and motor control. *Applied artificial intelligence* 13: 343-382
- Rossen, Brent, Cendan, Juan, Lok, Benjamin (2010). "Using Virtual Humans to Bootstrap the Creation of Other Virtual Humans". Short Paper. *10th International Conference on Intelligent Virtual Agents (IVA)*, Philadelphia, PA, Sept. 20-22, 2010.
- Rossen, B., Lind, D.S., Lok, B.: Human-centered Distributed Conversational Modeling: Efficient Modeling of Robust Virtual Human Conversations. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) *IVA 2009*. LNCS, vol. 5773, pp. 474–481.

Springer, Heidelberg (2009)

Rossen, B., SP Case link for interaction at http://vpf/cise.ufl.edu/VirtualPeopleFactory/publicOverview.php?u=3535&script_id=424a on February 13, 2011/

Sanders, G. A. and J Scholtz. 1998. Measurement and evaluation in embodied conversational characters. *In Proceedings of the Workshop on Embodied Conversational Characters* (WECC 98) (Tahoe City, Calif.), 114-118, 85-86. (Pagination due to proceedings printing error.)

Sims, M Edward. Reusable, lifelike virtual humans for mentoring and role-playing. *Computers & Education* Volume 49, Issue 1, August 2007, Pages 75-92
Web3D Technologies in Learning, Education and Training

Sixth Sense Technology, Google Image Search

Smith, Roger. 2009. Game Technology in Medical Education. An inquiry into the Effectiveness of New Tools 1st ed.

Standardized Patient Program, MEDED. (2004) Retrieved July 22nd, 2009 from <http://mededonline.usc.edu/sp.html>

Stephan, W., and W.E. Beane. 1978. The effects of belief similarity and ethnicity on liking and attributions for performance. *Revista Interamericana de Psicologica* 12:153-159

Tamblyn, M Robin. (1991). Sources of unreliability and bias in standardized-patient rating. *Teaching and Learning in Medicine, Volume 3, Issue 2 1991* , pages 74 - 85

Tamblyn, R.M., Klass, D.J., Schnabl, G.K., et al.: The accuracy of standardized patient presentation. *Medical Education* 25(2), 100–109 (2009)

Ti, L., Fun-Gee, C., Gee-Mei, T., Wah-Tze, T., Tan, J., Liang, S., et al. (2009). Experiential learning improves the learning and retention of endotracheal intubation. *Medical Education*, 43(7), 654-660. doi:10.1111/j.1365-2923.2009.03399.x.

Turner, M., Simon, S., Facemyer, K., Newhall, L., & Veach, T. (2006). Web-based learning versus standardized patients for teaching clinical diagnosis: a randomized, controlled, crossover trial. *Teaching And Learning In Medicine*, 18(3), 208-214. Retrieved from MEDLINE database.
<http://ezproxy.lib.ucf.edu/login?URL=http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=16776607&site=ehost-live>

US Army, (2010). A Request for Information. Three Dimensional and Holographic Display Systems Technology Roadmap and Current Markey Survey. Medical Simulation Applications.https://www.fbo.gov/index?s=opportunity&mode=form&id=3c8707e2c17d3ee4a5b84bfd70a2f311&tab=core&_cview=0

Wallace, Peggy (2007). Coaching Standardized Patients: For Use in the Assessment of Clinical Competence. New York, Springer Publishing Company

Wessel, J., Williams, R., Finch, E., et al.: Reliability and validity of an objective structured clinical examination for physical therapy students. *Journal of Allied Health* 32(4), 266–269 (2003).

Woodward, J., Carnine, D., & Gersten, R. (1988). Teaching Problem Solving Through Computer Simulations. *American Educational Research Journal*, 25(1), 72-86.

Wordnet Homepage. Available at: wordnetweb.princeton.edu/perl/webwn. Retrieved on December 18, 2010.