Annual Review of Cybertherapy and Telemedicine

Volume 5 Year 2007 ISSN: 1554-8716

Advanced Technologies in the Behavioral Social and Neurosciences

Editors:
Brenda K.Wiederhold, PhD, MBA, BCIA
Giuseppe Riva, PhD, MS, MA
Stéphane Bouchard, PhD

Interactive Media Institute

A Virtual Human Agent for Training Novice Therapist Clinical Interviewing Skills

Patrick Kenny, Albert A. Rizzo, Thomas D. Parsons, Jonathan Gratch & William Swartout
University of Southern California, Institute for Creative Technologies
kenny@ict.usc.edu

Abstract

Virtual Reality (VR) is rapidly evolving into a pragmatically usable technology for mental health (MH) applications. Over the last five years, the technology for creating virtual humans (VHs) has evolved to the point where they are no longer regarded as simple background characters, but rather can serve a functional interactional role. Our current project involves the construction of a natural language-capable virtual client named "Justin," which derived from a military negotiation training tool into a virtual therapy patient for training novice clinicians the art of clinical interviewing with a resistant client. Justin portrays a 16-year old male with a conduct disorder who is being forced to participate in therapy by his family. The system uses a sophisticated natural language interface that allows novice clinicians to practice asking interview questions in an effort to create a positive therapeutic alliance with this very challenging virtual client. Herein we proffer a description of our iterative design process and outline our long term vision.

Introduction

Virtual Reality (VR) is rapidly evolving into a pragmatically usable technology for mental health (MH) applications. As the underlying enabling technologies continue to evolve and allow us to design useful and usable "structural" clinical virtual environments (VE), the next important challenge will involve "populating" these environments with virtual representations of humans. Over the last five years, the technology for creating virtual humans (VHs) has evolved to the point where they are no longer regarded as simple background characters, but rather can serve a functional interactional role (Swartout et al., 2006; Gratch et al., 2002). This will be vital to create MH training tools that leverage the use of VHs for applications that require human-to-human interaction and communication. This would open up possibilities for clinical applications that address interviewing skills, diagnostic assessment and therapy training.

Virtual patients (VPs) are virtual interactive agents who are trained to simulate a particular clinical presentation of a patient with a high degree of consistency and realism. VPs have commonly been used to teach bedside competencies of bioethics, basic patient communication and history taking, and clinical decision making (Dickerson et al., 2005; Fleetwood et al., 2000; Johnsen et al., 2005; 2006; McGee et al., 1998; Stevens et al., 2005). VPs can provide valid, reliable, and applicable representations of live patients (Triola et al., 2006). For example, in Lok's application (Raij et al., 2005, 2006), instead of using the costly and labor intensive approach of hiring professional patients for novice medical students to practice on, they constructed a VE to represent an examination room where a VP could be interviewed verbally using the *Dragon Naturally Speaking* speech recognition software. The goal in this application was to determine, via clinical interview, whether the VPs ailment was due to appendicitis.

Research into the use of VPs in psychotherapy training is very limited (Kiss et al., 2003; Frank et al., 2002). Beutler and Harwood (2004) describe the development of a VR system for training in psychotherapy and summarize training-relevant research findings. We could not find reference to any other use of VPs in a psychotherapy course to date, despite online searches through MEDLINE, Ovid, and the psychotherapy literature.

The USC Institute for Creative Technologies has been conducting similar VH research as part of its primary mission over the last seven years to create highly interactive artificially intelligent agents to be used for VR military leadership and negotiation training (Rickel et al., 2001). This VH effort is built on prior work in the areas of embodied conversational agents (Cassell et al., 1998) and animated pedagogical agents (Johnson et al., 2000), but integrates a broader set of capabilities than any prior work. For the types of training scenarios we are targeting, the VHs must integrate three broad influences on their behavior: they must perceive and act in a 3D virtual world, they must engage in face-to-face spoken dialogues with people and other VHs in such worlds, and they must exhibit human-like non-verbal behavior and emotions. Traditional work on VH in the computer graphics and games community has focused on graphical look, perception and action in 3D worlds, but largely ignored dialogue and emotions.

Our current project involves the construction of a natural language-capable VH agent, named "Justin." The clinical attributes of Justin were developed to mimic a conduct disorder profile as found in the Diagnostic and Statistical Manual of Mental Disorders Text Revision (DSM-IV-TR; APA, 2000). The VP system is based on our existing VH architecture (Swartout et al. 2006). The VH architecture was derived from a military negotiation training tool into a virtual therapy patient for training novice clinicians the art of clinical interviewing with a resistant client. Justin portrays a 16-year old male with a conduct disorder who is being forced to participate in therapy by his family. The system uses a sophisticated natural language interface (Leuski et al., 2006) that allows novice clinicians to practice asking interview questions in an effort to create a positive therapeutic alliance with this very challenging virtual client.

Methods

The project involved the development of a VP, "Justin", as well as the clinical virtual environment (VE) in which the trainees learn interviewing techniques and knowledge of signs and symptoms of conduct disorder. The environment was modeled after a typical clinician's office and was meant to represent a place that would make the patient feel at home. The model representing "Justin" was meant to represent a typical 16 year old boy with jeans and a baseball hat. (See Image in Figure 1). The VP is capable of responding to clinical questions with verbal behavior and non-verbal human-like emotions and body postures that one might typically observe in a client with this disorder. Several role-playing scenarios between humans were conducted and recorded to gather the dialog and non-verbal behavior and voice of the VP. The preliminary goal is to use the VP to teach diagnostic skills training specifically for conduct disorder, as described below. The eventual goal is to have the VP be utilized in individual trainee interviews, small group and classroom settings.

The structure of the individual trainee interview targets an interview of 30 minutes duration during which there may be pauses for guidance from live supervisors who are in attendance,

however, the interview may progress successfully without interruption guided by verbal and emotional interactions between VP and trainee doctor. At the end of the 30 minute period, the trainee is expected to be able to make the diagnosis, having elicited the appropriate signs and symptoms, and conducted a sensitive and effective interview. If the diagnosis or the interview process is not successful within the expected time allotted, the trainee will need additional practice to be considered competent in the clinical diagnosis of conduct disorder.

The VP, "Justin", and the clinicians' activity and verbal behavior can be recorded during the interview and the entire 30 minute interview process may be re-played for review, critique and commentary by child and adolescent psychiatry attendings, as a teaching tool for other residents, or for groups of medical students learning about conduct disorder.

A competency-based set of questions and a scoring system is being developed in order to help child and adolescent psychiatry residents, psychiatry residents, and medical students utilize the recorded versions of the interviews as learner-centered educational tools. A library of various clinician-recorded interviews can be available to be viewed by students who have not actually participated in the interview, and can be interrupted at various points to assess the viewer's own diagnostic skill and response to the interview as it progresses. An example of a scoring system may include demonstration of knowledge of the core features of the disorder (in this case, conduct disorder) and measures of perceived strength and weakness of the interviewer's rapport with the VP. Diagnostic questions can prompt the viewer with respect to missing diagnostic symptoms or additional information as needed.

Conduct Disorder Diagnosis Skills Training

Teaching interviewing skills with VHs and VPs is still a young discipline. There are no standard methods and metrics. The larger problem of teaching general interviewing skills is even vaguer as there are many techniques and it is not well understood how to properly implement those with a VP. To alleviate this problem we are concentrating on teaching skills required to diagnose a particular disorder, in this case conduct disorder. Our goal is to obtain objective data from an initial intake interview. This will be accomplished by evaluating the questions asked by the trainee to the VP and the corresponding answers. The trainee's interview questions should be guided by the need to determine if the patient is internalizing or externalizing their behaviors and for eliciting information regarding the four general symptom categories prevalent in conduct disorder:

- Aggressive behavior – e.g. fighting, bullying, being cruel to others or animals
- Destructive behavior – e.g. arson, vandalism
- Deceitful behavior – e.g. repeated lying, shoplifting, breaking into homes or cars
- Violation of rules – e.g. running away, engaging in non appropriate behavior for age

The VP system is designed to provide answers to questions that target each of these categories and will respond to a variety of questions pertinent to these areas. Some responses by the VP may be on target, off target, involve "brush away" responses, and in some cases, they may be irrelevant replies. The probability of a specific response being emitted is rated to the question asked. For example if the trainee asks: "How are things going at home" or "Are you having any problems at home" or "How are things going?". The system will respond with "My parents think I messed up." Further questions will lead to finding out that the patient has been running away.

This will lead to marking one of the above categories true for the diagnosis in the trainees' interview. In order for the trainee to pass it will require responses in all of the categories. The total set of questions and responses are extracted from role playing exercises, initial subject testing, interviews with doctors and common sense for specific responses. In total a question set would consist of over 100-200 lines of text. The matching of questions to responses is a manual process with automated learning techniques to generate probability values (Leuski et al, 2006). Through an iterative process a good training set is developed.

Virtual Justin: Architecture and Integration

The VP system is based on the VH architecture developed at The Institute for Creative Technologies (Gratch et al., 2002, Swartout et al., 2006). The general architecture supports a wide range of VHs from simple question/answering to more complex ones that contain cognitive and emotional models with goal oriented behavior. The architecture is a modular distributed system with many components that communicate by message passing. Interaction with the system works as follows (See Figure 1): the trainee talks into a microphone which records the audio signal that is sent to a speech recognition engine. The speech engine converts the signal into text. The text is then sent to a statistical natural language system that matches the input text to a question/answer pair which selects an answer. The answer is sent to a non-verbal module which applies rules to create the appropriate gestures and behaviors. A procedural animation system then synchronizes the gestures, speech and lip synching and plays a pre-recorded or generated voice of the input text for the character for final output to the screen. The user then listens to the response and asks more questions to the character.

The major components in the system are: (See Figure 1)

- *Speech recognition*: Digitizes the trainee's speech and produces as output a string of words. The speech recognition engine used is SONIC from the University of Colorado, Boulder (Pellom, 2001). The language and voice models that SONIC uses are customized by us for the domain of interest (Sethy et al., 2005).
- *Natural Language Understanding and Dialog Management*: Parses the word string produced by speech recognition and forms an internal semantic representation. This representation selects an appropriate response from a set of statistically ranked questions (Leuski et al., 2006).
- *Non-verbal behavior generation*: Takes the response output string and applies a set of rules to select gestures, postures and gazes for the virtual character (Lee et al., 2006).
- *Intelligent Agent*: Reasons about plans and generates actions. Complex agents can be created with the Soar Cognitive architecture, or simple agents can be created with a finite state machine. Soar is a rule based system that contains goal directed behavior, short and long term memory, a reasoning mechanism, and can deal with input and output in real-time. The complex agents contain task models to reason about what actions to take. They also include a dialogue manager and a model of emotions.
- *SmartBody*: Synchronizes speech output with gestures and other non-verbal behavior to perform character control over the characters in the virtual environment.

- *Rhetorical*: Performs speech synthesis from the text generated by the agent. Or alternatively a set of pre-recorded voice strings can be used as the speech.
- *Unreal Tournament*: The current underlying graphics engine used for the Virtual Environment, we are exploring lower cost Open Source engines such as Ogre and Panda3D.

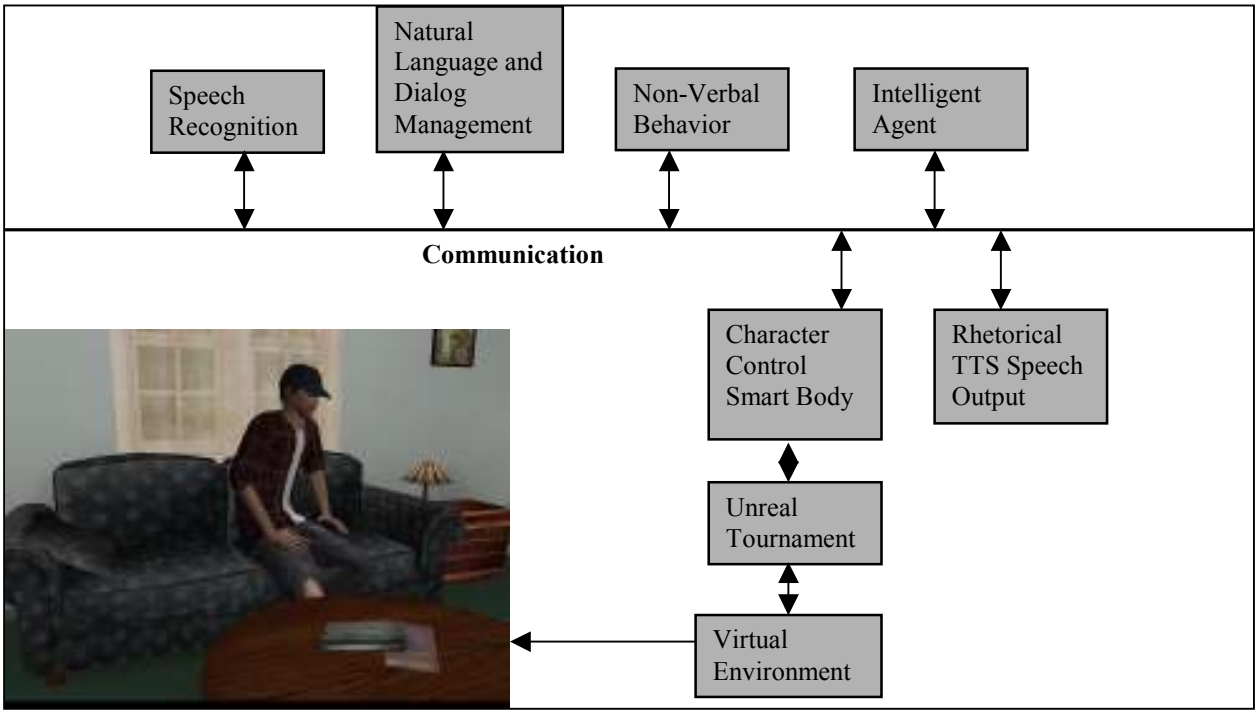

Figure 1. Virtual Human Architecture Diagram

Virtual Justin: Case Study

The VP is modeled after a conduct disorder case found in the *DSM-IV Casebook* (Spitzer et al., 1994). The virtual Justin is a 16 year old patient residing in the psychiatric unit of a general hospital in the wealthy Los Angeles suburb in which he lives. He had slashed his wrist with a knife, severing nerves and tendons in his left hand, and drifted in and out of consciousness during the night, finally calling a friend's mother for help in the morning. Justin is the son of a Vietnamese mother and an American serviceman. He lived with his mother in Saigon until he was two, when he came to the United States to be adopted by an American family through an agency specializing in adoption of Vietnamese children. He was apparently abused (burned and beaten) in this family, removed to a foster home for a brief period, and, at age four, placed with his current adoptive parents. Although Justin initially did well in his new surroundings and got along well with his friends, his relationship with his parents was turbulent, and they describe him as difficult.

Shortly after starting junior high, Justin began spending the majority of his time with a group of students who were many times truant from school. He also began adopting the behaviors of these students: smoking marijuana, shoplifting beer from the local supermarket, and belittling the values of parents and teachers. During this time his academics suffered and he was placed on

academic probation. His parents' attempts to redirect his attitude and relations were met with hostility. Justin got into trouble for fighting in school, blowing up mailboxes with firecrackers, and throwing darts at the neighbor's cat.

During this time, Justin's parents separated and he decided to stay with his father. The disruption of his parent's separation was accompanied by increased "acting out" behaviors. He and his friends were arrested for "borrowing" a car to go joyriding. Additionally, his truancy escalated to the point that he was absent more days than he was in school. While truant from school he and his friends would actively use drugs—mostly LSD, mescaline, glue, and marijuana.

Virtual Justin: DSM-IV-TR Diagnosis

Differential Diagnosis

Although Justin's ticket of admission to a psychiatric hospital seemed to be a suicide attempt, he later tells us that it was a clever way of avoiding being arrested. Whether or not he is also depressed and did intend to kill himself, there does not seem to be evidence of a full Major Depressive Episode that would justify that diagnosis. Further, we do not have enough information to make a positive diagnosis of Dysthymic Disorder, although his feeling that life is pointless suggests that this is likely (see DSM-IV-TR for major diagnostic criteria; APA, 2000).

Diagnosis

Justin has a history of a chronic pattern of antisocial behavior in which the basic rights of others and age-appropriate societal norms are violated. He has stolen, been truant, broken into someone's car, been cruel to animals, and initiated physical fights. These behaviors warrant the diagnosis of Conduct Disorder of Adolescent-Onset Type with Moderate Severity. It is understood to be of Adolescent-Onset Type because onset of conduct disturbance was in adolescence. The label of Moderate Severity is applied because Justin does not cause considerable harm to others or engage in extensive vandalism or theft (see DSM-IV-TR for major diagnostic criteria; APA, 2000).

Additional Diagnostic Considerations:

It is also important to note that Justin certainly has abused various drugs. We do not have sufficient information to know whether he was ever dependent on drugs. Given that his suicide attempt was related to his use of LSD, it appears that his use of hallucinogens is certainly having negative consequences, justifying the diagnosis of LSD Abuse (see DSM-IV-TR for major diagnostic criteria; APA, 2000). Similarly, the extensive use of marijuana, leading to repeated absences from school, would justify the diagnosis of Cannabis Abuse (see DSM-IV-TR for major diagnostic criteria; APA, 2000). We realize that careful questioning would probably reveal abuse of other drugs.

The diagnosis that best accounts for the current admission is Adjustment Disorder, With Depressed Mood because his depressed mood and suicide attempt seem to have been triggered

by the stress of what he believed was an impending arrest (see DSM-IV-TR for major diagnostic criteria; APA, 2000).

Conclusions

Herein we described the initial plan for developing and using a virtual patient that will be used for examining initial user data from a sample of psychiatric residents and psychology graduate students. This is an initial pilot study that will serve as the basis for a longer term research vision, that of creating a comprehensive Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR (APA, 2000) diagnostic trainer having virtual humans that are modeled after the symptoms and behaviors that are specified in each diagnostic category. Our long term plan is to also integrate an automated tutor agent that can appear at the end of training interview session to review the interview dialog with the trainee and provide feedback as to how a line of questioning may have properly gathered relevant assessment information or in a case where the effort became side tracked or convoluted.

We believe that virtual patients will play an important role in the future of psychotherapy education for psychiatry residents and psychology trainees. The use of virtual patients could be implemented in several ways. For example, virtual patients could be developed to recognize the essential features and common pitfalls of an initial psychotherapy interview so that they could give more specific, relevant, and reliable verbal feedback to the residents involved. In addition, the use of virtual patients illustrating common problems such as acting out, transference, intrusive questions, or seductive behavior would allow residents to have an experience with these anxiety-provoking situations in a simulated setting before they occur in their practice. Finally, performance in virtual patient scenarios could be used as an additional source of data for the assessment of resident competency in the psychotherapy domain.

Future work with the system would include: 1) addition of a camera for more user input into the system; 2) addition of more personality to the character; 3) maintain more conversation and world state about what is being discussed in the interview; 4) addition of intonation, prosody, and affect to the speech output (e.g given that persons with Conduct Disorder are prone to anger, it may be appropriate to have him shout his responses when the same question is asked several times); 5) addition of tools to build characters that have several different behavior problems, the dialog they would use and the non-verbal behavior they would manifest. For interview training a series of small vignettes that guide one particular interviewing technique such as reflective listening and following would be useful.

This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders – fourth edition – text revision.* Washington, D.C.: American Psychiatric Association.

Beutler, L.E., & Harwood, T.M. (2004). Virtual reality in psychotherapy training. *Journal of Clinical Psychology, 60*, 317–330.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., and Yan, H. (1998). "An Architecture for Embodied Conversational Characters.", *Proceedings of the First Workshop on Embodied Conversational Characters*, October 12-15, Tahoe City, California.

Dickerson, R., Johnsen, K., Raij, A., Lok, B., Hernandez, J., & Stevens, A. (2005). Evaluating a script-based approach for simulating patient-doctor interaction. *Proceedings of the International Conference of Human-Computer Interface Advances for Modeling and Simulation.*

Fleetwood, J., Vaught, W., Feldman, D., Gracely, E., Kassutto, Z., & Novack, D. (2000). MedEthEx online: a computer-based learning program in medical ethics and communication skills. *Teach Learn Med., Spring;12*, 96–104.

Frank, G., Guinn, C., Hubal, R., Pope, P., Stanford, M., & Lamm-Weisel, D. (2002). JUSTTALK: An application of responsive virtual human technology. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference, USA.*

Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., Petajan, E., (2002) Creating Interactive Virtual Humans: Some Assembly Required, *IEEE Intelligent Systems,* July/August, 54-63

Johnsen, K., Dickerson, R., Raij, A., Harrison, C., Lok, B., Stevens, A., et al. (2006). Evolving an immersive medical communication skills trainer. *Presence: Teleoperators and Virtual Environments, 15(1)*, 33-46.

Johnsen, K., Dickerson, R., Raij, A., Lok, B., Jackson, J., Shin, M., et al. (2005). Experiences in using immersive virtual characters to educate medical communication skills. *Proceedings of the IEEE Conference on Virtual Reality.*

Johnson, L., Rickel. J., Lester, J., (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, 11, 47-48.

Kallmann, M., Marsella, S. (2005). Hierarchical motion controllers for real-time autonomous virtual humans. *Proceedings of the 5th International working conference on Intelligent Virtual Agents (IVA),* 243–265, Kos, Greece.

Kiss, B., Szijarto, G., Benedek, B., Simon, L., Csukly, G., & Takacs, B. (2003). CyberTherapy: Applications of virtual reality and digital humans in clinical psychology. *Proceedings of the 2nd International Conference on Computer Animation & Geometric Modeling,* Hungary.

Lee, J., Marsella, S. (2006). Nonverbal Behavior Generator for Embodied Conversational Agents. *6th International Conference on Intelligent Virtual Agents,* pp 243-255, Marina del Rey, CA.

Leuski, A., Patel, R., Traum, D., Kennedy B., (2006) Building effective question answering characters. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, Sydney, Australia.

McGee, J.B., Neill, J., Goldman, L., Casey, E. (1998). Using multimedia virtual patients to enhance the clinical curriculum for medical students. *Medinfo, 9(Part 2)*, 732–735.

Pellom, B. (2001). Sonic: The University of Colorado continuous speech recognizer. *Technical Report TR-CSLR-2001-01*, University of Colorado, Boulder, CO

Raij, A., Johnsen, K., Dickerson, R., Lok, B., Cohen, M., Duerson, M., Pauly, R., Stevens, A., Wagner, P., Lind D., (2006). Comparing Interpersonal Interactions with a Virtual Human

to those with a Real Human. *IEEE Transactions on Visualization and Computer Graphics*.

Raij, A., Johnsen, K., Dickerson, R., Lok, B,. Cohen, M., Stevens, A., Bernard, T., Oxendine, C., Wagner, P., Lind, S., (2005). "Interpersonal Scenarios: Virtual Real?" In *Proceedings of IEEE Virtual Reality 2006*, 59-66.

Rickel, J., Gratch, J., Hill, R., Marsella, S., Swartout, W., (2001). Steve Goes to Bosnia: Towards a New Generation of Virtual Humans for Interactive Experiences. *In AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*, Stanford University, CA, March 2001

Sethy, A., Georgiou, P., Narayanan, S. (2005). Building topic specific language models from webdata using competitive models. In: Proceedings of EUROSPEECH, Lisbon, Portugal

Spitzer, R.L., et. al. (1994). *DSM-IV casebook: A learning companion to the diagnostic and statistical manual of mental disorders*. Washington, DC: American Psychiatric Press.

Stevens, A., Hernandex, J., Johnsen, K., et al. (2005). The use of virtual patients to teach medical students communication skills. *The Association for Surgical Education Annual Meeting*, April 7–10; New York, NY.

Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel J., Traum, D., (2006). Toward Virtual Humans, *AI Magazine,* v.27(1)

Triola, M., Feldman, H., Kalet, A.L., Zabar, S., Kachur, E.K., Gillespie, C., et al. A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal of General Internal Medicine, 21(5)*, 424-429, May, (2006)