Virtual Child Witness: Effects of single and multiple use on performance with Novice and Expert cohorts in a structured virtual human interview

Thomas B. Talbot, MD, Thomas D. Lyon, JD, PhD, Albert Rizzo, PhD, Bruce John
University of Southern California
Los Angeles, CA
talbot@ict.usc.edu, tlyon@law.usc.edu, rizzo@ict.usc.edu, brucejoh@usc.edu

ABSTRACT

Virtual human avatars can be used to train and assess a myriad of complex skills, such as interviewing, interpersonal, and clinical skills, in a safe environment that provides consistency, reduced cost, greater accessibility, and objective feedback. We created a structured virtual human interview which consisted of a conversational avatar that interacts verbally in response to on screen question choices. Our prototype was a forensic interview simulation called Virtual Child Witness (VCW). VCW provides a content-rich interview in response to open-ended questions and is designed to assess user's interviewing strategy. In a quasi-experimental design, we evaluated 222 subjects to determine if the system could discriminate between Experts ($M = .713$, $SD = .29961$) and Novices ($M = .373$, $SD = .27547$), $t(230) = 9.002$, $p < .001$, $d = 1.18$, indicating easy discrimination between cohorts. A subset of 92 subjects completed the simulation multiple times; multiple use was associated with much higher performance ($M = .727$, $SD = .23669$) than single use ($M = .489$, $SD = .33688$), $t(88.847) = 5.491$, $p < .001$, $d = .92$. Our results demonstrate that multiple use improved performance and that differing skill level is measured by the VCW system. The gamut of question topics and the question-dependent branching architecture of the conversation made a practice effect unlikely, as compared to subjects interpreting feedback data and adapting their interview strategy. Choice-based virtual interviews have the advantage of using inexpensive technology that does not require the language processing artificial intelligence of more complex systems. The system tracked increases in performance with a group of subject Novices that was near equivalent to a single performance from a subject-matter Expert group. This study demonstrates that structured virtual encounters can provide valid assessments and that intentional practice can improve performance. This successful design can serve as a model for virtual interview training in other professional domains such as military interrogation, standardized medical patients, and national security applications.

ABOUT THE AUTHORS

Thomas B. Talbot, MD is an Associate Professor of Medical Education at the University of Southern California Keck School of Medicine & Principal Medical Expert at the USC Institute for Creative Technologies. Dr. Talbot has a 19-year history creating medical games and simulations. Much of his current work involves virtual human conversational agent training for the military and the medical community.

Thomas D. Lyon, JD, PhD is the Judge Edward J. and Ruey L. Guirado Chair in Law and Psychology at the University of Southern California Gould School of Law. His research involves increasing the productivity of and minimizing errors in child witnesses' reports.

Albert "Skip" Rizzo, PhD is Director of Medical Virtual Reality at USC Institute for Creative Technologies. He is a pioneer in the use of VR for clinical applications including VR exposure therapy for Post-Traumatic Stress Disorder.

Bruce John is currently a freelance research support specialist who, at ICT, served as Research Coordinator on the VCW effort. He, among many crucial roles, designed the VCW prototype and conducted data analysis.

Virtual Child Witness: Effects of single and multiple use on performance with Novice and Expert cohorts in a structured virtual human interview

Thomas B. Talbot, MD, Thomas D. Lyon, JD, PhD, Albert Rizzo, PhD, Bruce John
University of Southern California
Los Angeles, CA
talbot@ict.usc.edu, tlyon@law.usc.edu, rizzo@ict.usc.edu, brucejoh@usc.edu

INTRODUCTION

We are in the midst of a pedagogical revolution in fields where difficult to teach skills need to be imparted to new generations of doctors, clinicians, investigative interviewers, and military personnel. Such training poses potentially high-risks when engaged *in vivo*, and so trainers must turn to alternative methods for preparing these skills effectively, while at the same time minimizing the risks to actual patients and/or witnesses as well as the trainees themselves.

The use of virtual characters is one approach to training these skills in a safe environment that can provide unparalleled consistency, reduced cost, greater accessibility, and a standardized feedback system. The Virtual Child Witness (VCW) project provides a safe, digital environment for new and veteran investigative interviewers to practice and hone their interviewing skills by talking to a virtual child. This allows investigative interviewers to make mistakes with the computer that, if they were to make with an actual maltreated child witness, could have devastating repercussions *in vivo*.

Research under consideration follows: Does using the VCW system multiple times improve performance in the system? (RQ1)? Does the interview skill-level of the participant influence performance in the system? (RQ2) Does the skill-level of the interviewer influence performance improvements when using the system multiple times? (RQ3).

BACKGROUND

One early goal of the VCW project was to determine if Novice and Expert level interviewers were performing differently with the virtual child. Our hypothesis was that the way in which the program is designed, and with the performance metrics it gathers, the program should be able to differentiate between different skill-levels of investigative interviewers. In a pilot study, an earlier version of the Virtual Child Witness (VCW) program showed significant differences between Expert and Novice level interviewers. Forty-five participants used the first version, 22 in the Expert group recruited from the USC Law School, and 23 in the Novice group recruited from the USC Institute for Creative Technologies. An independent-samples t-test showed that the Expert group ($M = .860$, $SD = .151$) asked a significantly higher percentage of open-ended questions compared to the Novice group ($M = .477$, $SD = .284$), $t(24,748) = 5.150$, $p < .001$, $d = 1.68$, with a 95% CI; these analyses also showed significant differences between the groups with regards to the other question types (John, Talbot, Lyon, Rizzo, & Buckwalter, 2013).

This early version of the program proved very promising, and through internal funding sources, the Virtual Child Witness 2.0 (VCW) program was recreated in the University of Southern California's Institute for Creative Technologies' SimCoach platform (Rizzo et al., 2010). The SimCoach platform was utilized because it provided four distinct advantages over the previous engine used to drive the system; 1) The ability to deploy to, and be accessible from, a web-browser; 2) Greater flexibility in the creation and arranging of interaction content; 3) Increased capability to provide after-action review feedback to users; and 4) Improved visual fidelity of the virtual child character.

Choice-based virtual human encounters can be employed in a variety of educational scenarios. We have employed such scenarios for military leadership training, interpersonal counseling, and medical interview training (Talbot, 2016). The purpose of the VCW interaction is to simulate important early steps in a child forensic interview (Lyon, 2005). Such interviews are extremely sensitive and it is important to ensure that evidence obtained from a child witness is accurate and admissible in court. (Fisher, 1999) There is international consensus that interviewers should

use open-ended questions, which have been shown to encourage more elaborate, accurate and comparably unsolicited information (Lipton, 1977). Unfortunately, even professionals have difficulty adhering to open-ended questions. There is evidence that practice with trained interview surrogates, similar to standardized patients in medicine, is effective in providing the necessary training to adapt their interviewing style (Powell, 2008a)

VCW employs a virtual standardized child witness. In the tradition of effective training (Powell, 2008b), the child witness recounts the details of an innocuous event. In this case, the virtual child witness is asked about his last birthday party. In the simulation, the interviewer may choose from a variety of questions of the following types: open-ended (OE), Wh-questions such as what, when, where, why, etc. (-Wh), and yes/no questions (Y/N). The simulation prefers open-ended questions and responds to that question type with more elaborate responses and the unlocking of more conversation topics and answer dialogue.

Figure 1. Virtual Child Witness 2.0 Prototype. Possible questions are seen on the lower left. Additional topics to explore are on the lower right (currently greyed out). Character responses are verbal.

METHODS

We recruited subjects for an evaluation of 'Experts' and Novices in order to determine the VCW system's ability to discriminate between these groups. Novices were paid participants whom we recruited through Amazon Mechanical Turk™. Novices were screened out for a number of criteria, including experience as a child interviewer. Novice demographics matched the general population except that a higher percentage were college educated, they were more familiar with computer use and Asians were overrepresented (9%). The Expert group consisted of professionals attending a forensic interviewing seminar at USC. This group included law students, social workers, and law enforcement. Experts were invited to participate in VCW but were not required to do so. A fairly large number of Experts completed an initial screening with VCW as they started an interviewing course, with a minority completing a second VCW interview upon course completion.

All data from the SimCoach virtual child witness studies were inspected to determine whether or not they had participant ID numbers. A total of 711 cases were inspected (this is data that gets generated anytime someone uses the VCW system, e.g. studies, demos, and other testing). All data with a participant ID were retained, while any data that did not have a participant ID were removed, leaving a total of 581 remaining cases. Remaining data were inspected to determine if the participant ID number was present more than once. If the ID was present more than once, then all

data corresponding to that ID were retained. Any data with only a single set of data corresponding to the ID were removed leaving a total of 92 participants who had gone through the interaction more than once. The remaining data were again screened to ensure that at least one of the sessions attached to a participant ID had available performance data, 44 participants remained in the data set after this screening. Remaining cases were removed to ensure that a participant had only the most recent performance score in the dataset. The valid one-session cases were then merged back into the data, yielding a dataset with 44 participants with multiple sessions and 188 with single sessions (total N = 232). Of the multiple session participants, 19 were Novices and 25 were Experts. Of the single session participants, 103 were Novices and 85 were Experts. In our original studies it was intended that Novice subjects take the simulation once, at which point they were provided performance metrics (Figure 2) and instructed to complete a post-questionnaire which afterwards signified their completion of the study. It was suggested to the seminar attendees that they could take the VWC assessment at the beginning and end of the seminar. The 44 multiple use subjects completed the VCW simulation at least two times and had the opportunity to view the performance metrics after each try.

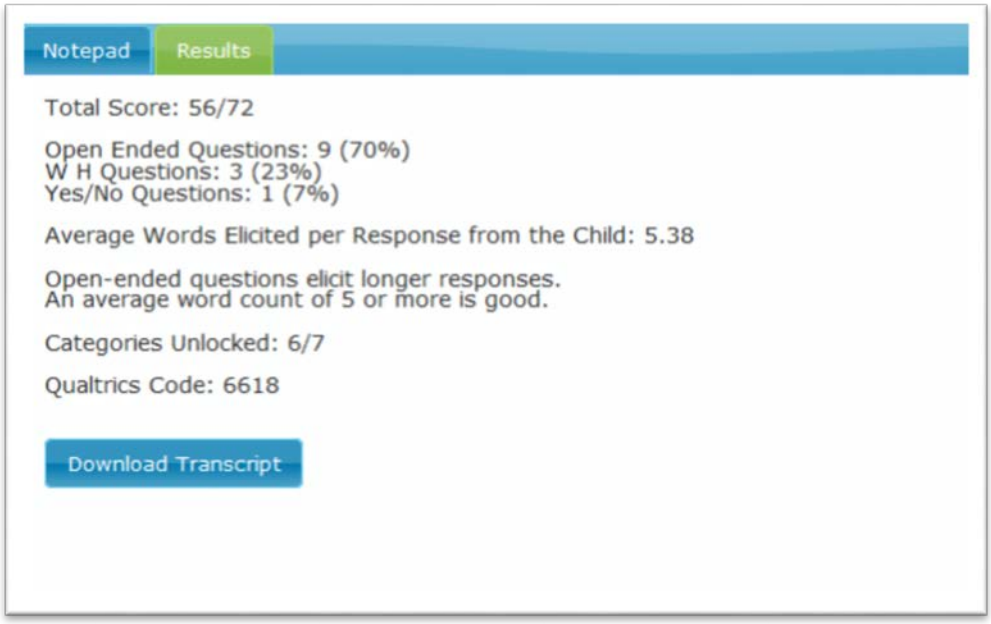

Figure 2. VCW Assessment Feedback.

Evaluating Subject Responses

For evaluation purposes, excellent user performance was reflected by open-ended questions (Lamb, Hershkowitz, Orbach, & Esplin, 2008) which yielded longer responses by the system and resulted in greater availability of follow up topics. Interviewer questions were classified into three types; open-ended (OE), Wh-questions (-Wh), and yes/no questions (Y/N). While measuring the frequency of the question types asked, as well as total percentage a question type was asked, a final composite score is also computed by assigning points to the different question types; OE questions give the most points, -Wh questions give 3.5 times fewer points than OE, and –Y/N questions give a single point which is half that of -Wh questions. The system also tracks the average number of words the child responds with after each question, with the child giving more detailed responses for OE questions, and slightly more detailed responses for –Wh questions as compared to Y/N questions. The last performance metric tracked by the system is the number of question categories unlocked by the participant, if the person asks more appropriately timed questions of the right question type then they will be able to explore more parts of the interaction that would otherwise get closed off by asking the wrong question (Table 1).

Table 1. VCW Performance Metrics

Performance Metrics			
Metric Name	**Min Value**	**Max Value**	**Interpretation**
Total Score	0	72	Higher↑ value means better performance
OE Percent	0.00	1.00	Higher↑ value means better performance
-Wh Percent	0.00	1.00	Lower↓ value means better performance
Y/N Percent	0.00	1.00	Lower↓ value means better performance
Average Response Word Count	.14	7	Higher↑ value means better performance
Question Categories Unlocked	0	7	Higher↑ value means better performance

RESULTS

Research Question 1 (RQ1) – Effect of VCW Multiple Use

Whether or not the participant had multiple sessions served as the independent variable, with system performance metrics serving as the dependent variables. Independent-samples T-tests indicated that those who engaged in multiple sessions with the program performed significantly better than those who only used the program once (see table 2 for T-Tests). All performance metrics were significant at the .05 level except –Wh percent and -Wh frequency.

Two of the largest effects were seen with OE question percentage and Y/N question percentage; the group of participants who used the program multiple times asked a significantly higher percentage of OE question types ($M = .727$) than the group of participants who used the program only once ($M = .489$, $t(88.847) = 5.491$, $p < .001$, $d = .92$.) The group of participants who used the program multiple times asked a significantly lower percentage of Y/N question types ($M = .1261$) than the group of participants who used the program only once ($M = .3289$), $t(118.532) = -6.24$, $p < .001$, $d = 1.05$.

The multi-session group had higher values on all performance metrics that were associated with higher values meaning better performance. Except in frequency of –Wh questions where they were relatively equal, the multi-session group also had lower values on all performance metrics that were associated with lower values meaning better performance (see Table 2 for group statistics and effect sizes). In the Expert group, these gains could be due either to multiple use, or experience gained from their interviewing seminar as this was a pre-post course activity. For the Novice group, the entire performance difference is due to multiple use.

Table 2. Effect of Multiple System Use

Group Statistics & Effect Sizes (RQ1)						
Metric Name	**Single vs. Multiple Sessions**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**	**Cohen's d (95% CI of d)**
Total Score	Single	188	35.22	21.807	1.590	0.91 (.571, 1.25)
	Multiple	44	52.68	18.538	2.795	
OE Percent	Single	188	.4891	.33688	.02457	0.92 (.581, 1.26)
	Multiple	44	.7270	.23669	.03568	
-Wh Percent	Single	188	.1820	.14168	.01033	Not Significant
	Multiple	44	.1468	.12809	.01931	
Y/N Percent	Single	188	.3289	.29382	.02143	1.05 (.703, 1.39)
	Multiple	44	.1261	.16201	.02442	
	Multiple	44	1.34	1.509	.227	
Average Response Word Count	Single	188	4.0861	1.96121	.14304	.95 (0.61, 1.29)
	Multiple	44	5.4802	1.33472	.20122	
Question Categories Unlocked	Single	188	4.03	1.965	.143	0.88 (.540, 1.22)
	Multiple	44	5.55	1.663	.251	

Research Question 2 (RQ2) - Effect of Interviewer Experience

As in the original VCW study, discussed in the introduction, recruitment for SimCoach VCW studies took place from two different populations. One group was made up of Novice-level interviewers and another group consisted of Expert-level interviewers. Whether or not the participant was an Expert or Novice served as the independent variable, with system performance metrics serving as the dependent variables. Independent-samples t-tests indicated that those who were Expert interviewers performed significantly better than those who were Novice interviewers. All performance metrics were significant at the .05 level.

The group of participants who were Experts asked a significantly higher percentage of OE question types (M = .713,) than the group of participants who were Novices (M = .373), t (230) = 9.002, $p < .001$, d = 1.18. The group of participants who were Experts asked a significantly lower percentage of Y/N question types (M = .1517) than the group of participants who were Novices (M = .4155), t (229.962) = -7.979, $p < .001$, d = 1.05.

The Expert group had higher values on all performance metrics that were associated with greater values being better in performance. The Expert group also had lower values on all performance metrics that were associated with lower values being better in performance (see Table 3 for group statistics and effect sizes).

Table 3. Interview Experience & VCW Performance

Group Statistics and Effect Sizes (RQ2)						
Metric Name	**Expert vs. Novice**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**	**Cohen's d (95% CI of d)**
Total Score	Novice	122	27.85	17.682	1.601	1.17 (0.89, 1.45)
	Expert	110	50.37	20.855	1.988	
OE Percent	Novice	122	.3731	.27547	.02494	1.18 (0.90, 1.46)
	Expert	110	.7130	.29961	.02857	
-Wh Percent	Novice	122	.2114	.14097	.01276	0.57 (0.30, 0.83)
	Expert	110	.1353	.12720	.01213	
Y/N Percent	Novice	122	.4155	.26643	.02412	1.05 (0.77, 1.32)
	Expert	110	.1517	.23704	.02260	
	Expert	110	1.26	1.629	.155	
Average Response Word Count	Novice	122	3.4168	1.61622	.14633	1.18 (0.90, 1.46)
	Expert	110	5.3861	1.72610	.16458	
Question Categories Unlocked	Novice	122	3.46	1.726	.156	1.02 (0.74, 1.29)
	Expert	110	5.27	1.847	.176	

Research Question 3 (RQ3) – Effect of Skill Level & Multiple Use

In order to assess the interaction of the main effects, multiple-usage and skill-level, a multivariate model was constructed with multiple-usage and skill level as independent variables and the system performance metrics as the dependent variables (Table 4).

Table 4. Independence of Interviewer Skill Group & Effects of Multiple Use

Full Group Statistics (RQ3)					
Metric Name	**Multi Session**	**Skill-Level**	**Mean**	**SD**	**N**
Total Score	Single	Novice	24.36	14.512	103
		Expert	48.38	21.979	85
		Total	35.22	21.807	188
	Multiple	Novice	46.79	21.447	19
		Expert	57.16	14.913	25
		Total	52.68	18.538	44
	Total	Novice	27.85	17.682	122
		Expert	50.37	20.855	110
		Total	38.53	22.271	232

The multivariate model confirmed that there was a significant interaction with the independent variables by themselves, multi-usage ($p = .001$) and skill-level ($p < .001$), and the performance metrics which served as the dependent variables. However, there was no significant interaction between the main effects ($p > .05$), giving assurance that both main effects are relatively independent of each other. It should be noted that the Expert group improvement was likely influenced by their interviewing course, so only changes in Novice group performance should be attributed to the effects of VCW use.

It did not matter whether a Novice used the system multiple times or not as performance gains were seen across the board with multiple uses of the system, Experts did better than Novices while multi-session users did better than single-session users. Though Novices improved to a larger degree, as measured, but also as limited, by the system, than Experts, both groups saw significant improvement gains (Figure 3). Expert gains can be attributed to multiple use and/or seminar completion whereas Novice gains are due exclusively to multiple use.

Summary of Research Results

Does using the system multiple times improve performance in the system? (RQ1) Based on the results of the analysis, there is evidence that those who used the system multiple times performed better than those who used the system only once. Does the interview skill-level of the participant influence performance in the system? (RQ2) Based on the results of the analysis, there is statistically significant evidence that those who are Expert interviewers perform better than those who are Novice-level interviewers, though the gap between Expert first time VCW users and Novice multiple users is slight. Does the skill-level of the interviewer influence performance improvements when using the system multiple times? (RQ3) Based on the results of the analysis, there is statistically significant evidence that both main effects of skill-level and multiple-usage are relatively independent of each other.

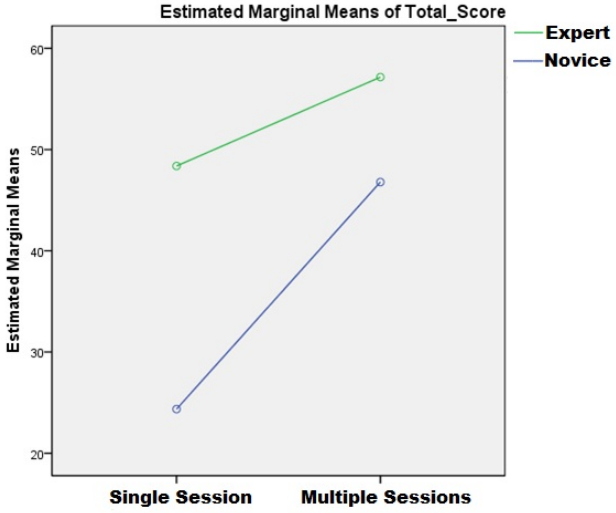

Figure 3. Total Scores After Single and Multiple VCW Use
Although the Novice score increase was more prominent, Novices started from a much lower performance baseline to achieve near-parity w/ first attempt Experts.

DISCUSSION

Structured virtual human interviews offer the benefit of a credible conversational interaction coupled with objective performance measurement and assessment. The gains we are attributing from system use are with the Novice group, which consisted of individuals who engaged in multi-session plan within the same day. This does not apply to the Expert group, where their repeat use occurred towards the completion of an interviewing course. The results did indicate that Experts did learn something from that interviewer training course, however.

The dynamic structure of the VCW was able to stimulate repeat-usage by users through its performance feedback system and the inherent uncertainty of what other information the virtual child might be able to provide. Both Expert and Novice multiple use subjects showed significantly higher performance. In fact, Novices performed at the pre-testing Expert level on the interaction. This is very important, as the Novice group received no interview training and depended exclusively on the VCW feedback system for improvement. Given that the Expert group came from a cohort that were self-enrolled in a child-interviewing seminar, the similarity between Novice repeat-performance scores and Expert first-time scores is impressive and encouraging as we suspect that Novice gains are a result of experiential learning. This is especially impressive when considering that the VCW feedback was very generic

(Figure 2) and did not provide feedback to individual responses. Feedback was limited to 1) a score, 2) question type percentages, 3) child response length, 4) # of categories unlocked and 5) the statements "Open-ended questions elicit longer responses. An average word count of 5 or more is good."

Repeat use of the system along with VCW randomization of response topics for each repeat performance exposed new conversation topics that rendered rote learning of responses to be an inefficient adaptation; learners were more likely to be applying new strategies, than memorizing responses or demonstrating a rehearsal effect. The idea of a conversational encounter, coupled with an assessment report and encounter repetition, may very well serve as a model for future virtual human-based interview training applications.

Study Limitations

Because this report and study data were the result of an opportunistic finding, this study has a number of important limitations. The individuals who committed to multiple VCW encounters from both groups did so on their own initiative; there is the potential for self-selection bias because of this. By taking only the first and last performance of participants, we are ignoring any measurable variability that might occur on a session-to-session basis. While refinements for determining the relation between size of improvement and number of simulation repetitions are a future goal for the system, the system also needs to be evaluated on a longer-term basis in a role that includes reinforcing knowledge and skills over the entirety of an interviewer's initial training and career. The Novice & Expert groups represent two independent variable measures because the conditions regarding system use are not the same (the Expert group had additional training from a course), hence only the Novice improvement effects can be attributed to use of the simulation.

Natural Language Random Access versus Structured Conversations

Another approach to conversational training is to develop a system that allows the user to say whatever they wish during an interview. Such a system is called a Natural Language Random Access (NLRA) virtual human encounter. We have developed several such systems, including USC Standard Patient, a virtual standardized patient (Talbot, 2016) as a freeware resource that is currently being used for Military training and have learned much from this approach. Although NLRA systems can be very effective and may seem tempting as an initial choice, they are costly and laborious to develop, they require significant Expertise to build a domain and they require a great volume of user testing to train the natural language artificial intelligence systems that they depend on (Talbot, Sagae, John, & Rizzo, 2012). Creating assessments for NLRA conversations is also a considerable pedagogical challenge.

In contrast, structured conversational encounters have numerous advantages in exchange for limiting the user's choices to a number of available options. They do not depend on natural language understanding technology, so they are certain to receive the user's intended input. Structured encounters are based on measurable choices that make assessment easy. User decision pathways can lead to defined outcomes and it is easier to build assessment feedback for these interactions. The conversations can appear more responsive as choices made can lead to conversation pathways with a markedly different virtual human demeanor. Structured encounters can also be produced at a very modest cost. Although we developed an online virtual human for the VCW system, it is possible to construct structured encounters within a web page or hyperlink-enabled PowerPoint presentation. The virtual human can be replaced with video clips of an actor. Thus, it is possible for a motivated educator to create structured conversational experiences inexpensively.

CONCLUSION

Our results indicate that deliberate repetitive practice along with feedback between rehearsals in a virtual forensic interview results in a strong, statistically-significant gain in performance. This is consistent with a growing body of literature regarding the efficacy of interview practice with virtual humans or avatars (Pompedda, Zappala & Santtila, 2015). The mechanism for the improvement using VCW is strategic adaptation on the part of the human interview and it is not a practice effect. The large VCW domain and topic randomization made a practice effect an unlikely contributor to this phenomenon. It is encouraging that random individuals can achieve comparable performance to professionals starting in a forensic interviewing seminar, though those same professionals did achieve a higher level of performance after seminar completion.

The feedback provided by VCW was fairly generic and the evaluation based on question type was simple. The next level of virtual conversation exploration would do well to evaluate more complex conversations and the inclusion of higher stakes question choices and more specific feedback. Our team is developing a new structured conversational system, called Select-A-Chat, that will allow for question by question feedback, unique metrics based on conversational parameters, educator scenario authoring and applicability for a variety of domains. Potential domains we have considered include patient interviews, interpersonal counseling, difficult conversations and negotiation training. Further research based on this study include leadership training, sexual harassment, and patient counseling applications for the Military.

Structured conversational training systems can provide definitive metrics for training organizations and they can be developed at a reasonable cost with low technical risk. They should be considered as a lower risk and cost alternative to NLRA interviewing systems when conversation-based simulation training is desired.

REFERENCES

Fisher, R. P. (1999). Probing knowledge. In D. Gopher & A. Koriat (Eds.), Attention and performance XVII (pp. 537–556). Cambridge, MA: The MIT Press

John, B. S., Talbot, T. B., Lyon, T., Rizzo, A., Buckwalter, J. G. (2013l). Training Effective Investigative Child Interviewing Skills Through Use of a Virtual Child: A Pilot Study Comparing Two Different Skill-Level Groups Performance with a Prototype of the Virtual Child Witness (VCW) Program. In 5th International Pediatric Simulation Symposia and Workshops; 2013 Apr 23 – Apr 25; New York City, United States. IPSSW; 2010 236 p.

Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). Tell me what happened: Structured investigative interviews of child victims and witnesses. John Wiley & Sons.

Lipton, J.P. (1977). On the psychology of eyewitness testimony. Journal of Applied Psychology, 62, 90–93.

Lyon, T.D. (2005). Ten step investigative interview, version 2, adaptation of the NICHD Investigative Interview Protocol. Accessed at https://works.bepress.com/thomaslyon/5/download/

Pompedda, F., Zappalà, A., & Santtila, P. (2015). Simulations of child sexual abuse interviews using avatars paired with feedback improves interview quality. Psychology, Crime & Law, 21(1), 28-52.

Powell, M.B, & Rebecca Wright, R. (2008a): Investigative Interviewers' Perceptions of the Value of Different Training Tasks on their Adherence to Open-Ended Questions with Children, Psychiatry, Psychology and Law, 15:2, 272-283 http://dx.doi.org/10.1080/13218710802014493

Powell, M.B., Fisher, R.P, & Hughes-Scholes, C.H. (2008b) The effect of using trained versus untrained adult respondents in simulated practice interviews about child abuse. Child Abuse & Neglect 32:1007-1016/

Rizzo, A., Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., Williams, J., et al. (2010). SimCoach: An intelligent virtual human system for providing healthcare information and support. In 2010 International Conference on Disability, Virtual Reality and Associated Technology; 2010 Aug 31 – Sept 2; Valparaíso, Chile. ICDVRAT; 2010 pp. 213 p.

Talbot, T.B. (2016) USC Standard Patient. Accessed at http://standardpatient.org

Talbot, T.B., Sagae, K., John, B., & Rizzo, A.A. (2012) Designing useful virtual standardized patient encounters. In *Interservice/Industry Training, Simulation and Education Conference* (*I/ITSEC*). (Vol. 2012, No. 1). National Training Systems Association.