Beyond Game Effectiveness Part II: A Qualitative Study of Multi-role Experiential Learning

Elaine M. Raybourn
Sandia National Laboratories*
Albuquerque, NM 87106
emraybo@sandia.gov

Nathan Fabian
Sandia National Laboratories*
Albuquerque, NM 87106
ndfabia@sandia.gov

Eilish Tucker
Sandia National Laboratories*
Albuquerque, NM 87106
emtucke@sandia.gov

Matthew Willis
Sandia National Laboratories*
Albuquerque, NM 87106
mwillis@sandia.gov

ABSTRACT

The present paper is the second in a series published at I/ITSEC that seeks to explain the efficacy of multi-role experiential learning employed to create engaging game-based training methods transitioned to the U.S. Army, U.S. Army Special Forces, Civil Affairs, and Psychological Operations teams. The first publication (I/ITSEC 2009) summarized findings from a quantitative study that investigated experiential learning in the multi-player, PC-based game module transitioned to PEO-STRI, DARWARS Ambush! NK (non-kinetic). The 2009 publication reported that participants of multi-role (Player and Reflective Observer/Evaluator) game-based training reported statistically significant learning and engagement. Additionally when the means of the two groups (Player and Reflective Observer/Evaluator) were compared, they were not statistically significantly different from each other. That is to say that both playing as well as observing/evaluating were engaging learning modalities. The Observer/Evaluator role was designed to provide an opportunity for real-time reflection and meta-cognitive learning during game play. Results indicated that this role was an engaging way to learn about communication, that participants learned something about cultural awareness, and that the skills they learned were helpful in problem solving and decision-making.

The present paper seeks to continue to understand what and how users of non-kinetic game-based missions learn by revisiting the 2009 quantitative study with further investigation such as stochastic player performance analysis using latent semantic analyses and graph visualizations. The results are applicable to First-Person game-based learning systems designed to enhance trainee intercultural communication, interpersonal skills, and adaptive thinking. In the full paper, we discuss results obtained from data collected from 78 research participants of diverse backgrounds who trained by engaging in tasks directly, as well as observing and evaluating peer performance in real-time. The goal is two-fold. One is to quantify and visualize detailed player performance data coming from game play transcription to give further understanding to the results in the 2009 I/ITSEC paper. The second is to develop a set of technologies from this quantification and visualization approach into a generalized application tool to be used to aid in future games' development of player/learner models and game adaptation algorithms.

Specifically, this paper addresses questions such as, "Are there significant differences in one's experience when an experiential learning task is observed first, and then performed by the same individual?" "Are there significant differences among groups participating in different roles in non-kinetic engagement training, especially when one role requires more active participation that the other?" "What is the impact of behavior modeling on learning in games?" In answering these questions the present paper reinforces the 2009 empirical study conclusion that contrary to current trends in military game development, experiential learning is enhanced by innovative training approaches designed to facilitate trainee mastery of reflective observation and abstract conceptualization as much as performance-based skills.

ABOUT THE AUTHORS

Dr. Elaine Raybourn is a Principal Member of the Technical Staff at Sandia National Laboratories where she leads computer game research in creativity, multi-role experiential learning, novel in-game assessment & debriefing techniques, and designing personalized training systems that stimulate intercultural communication competence and metacognitive agility. She has a Ph.D. in Intercultural Communication and has been designing computer games for

experiencing cultural dynamics for the past 11 years. Elaine's greatest passion involves enhancing intercultural communication and multicultural understanding for all through the design of experiential and collaborative virtual worlds, simulations, and games that promote communication, reflection, self-awareness, and intercultural adaptability. Since 2003 Elaine has led teams that have developed, deployed, or transitioned cross-cultural engagement game-based training for the U.S. Army Special Forces, DARPA, USMC, and US Army PEO-STRI. Her team's work was identified by the Defense Science Board 2006 Summer Study on *21st Century Strategic Technology Vectors* as "critical capabilities and enabling technologies for the 21st century that show promise." Elaine is a scientific advisory board member for the European Commission Integrated Project TARGET (Transformative, Adaptive, Responsive and EnGaging EnvironmenT), has been an ERCIM (European Consortium for Research in Informatics and Mathematics) fellow, and is currently on assignment to PEO STRI Games for Training, the Advanced Distributed Learning Co-Laboratory, and Defense Acquisition University. She is a recipient of the Department of the Army award for Patriotic Civilian Service.

Nathan Fabian is a Member of the Technical Staff for Sandia's Visualization and Analysis department. He has a Masters degree in Computer Science from the University of New Mexico with an emphasis on video game player behavior modeling with machine learning. His current work at Sandia is split between extreme scale scientific visualization and behavior-based informatics visualization.

Eilish Tucker is an Undergraduate Student Intern in Sandia's Cognitive Science Applications department. She is a third year student at Marquette University, studying Psychology and Cultural Anthropology.

Matthew Willis is a Research Assistant at Sandia's Cognitive Science Applications department, and has a M.A. in Health Communication with an emphasis in Technology. Matt has experience serving as a research assistant on research projects funded by the U.S. Army RDECOM, Office of the Secretary of Defense, and multiple National Institute of Health R1 grants. Matt co-teaches a specialized course with a Neurologist from the Albuquerque V.A. Spinal Cord Injury Center at the University of New Mexico entitled "The Web and our Brains"; the course explores the impacts of technology on human brain development. Also, Matt serves as the chair of the Research Committee for the Graduate and Professional Student Association at the University of New Mexico. Matt is currently a National Physical Science Consortium fellow.

Beyond Game Effectiveness Part II: A Qualitative Study of Multi-role Experiential Learning

Elaine M. Raybourn
Sandia National Laboratories*
Albuquerque, NM 87106
emraybo@sandia.gov

Nathan Fabian
Sandia National Laboratories*
Albuquerque, NM 87106
ndfabia@sandia.gov

Eilish Tucker
Sandia National Laboratories*
Albuquerque, NM 87106
emtucke@sandia.gov

Matthew Willis
Sandia National Laboratories*
Albuquerque, NM 87106
mwillis@sandia.gov

INTRODUCTION

Militaries have been utilizing game-based training for the past several years. Several game-based training studies have recently been made available (Brown, 2010; Raybourn, 2009; Belanich, et. al., 2004; Orvis, et. al., 2006; Beal, 2009, 2006; Rowan & Brown, 2008). With the exception of a study on game-based non-kinetic engagements (Raybourn, 2009) the findings from these studies primarily address kinetic training missions.

The present paper is the second in a series published at I/ITSEC that seeks to explain the efficacy of multi-role experiential learning employed to create engaging game-based training methods transitioned to the U.S. Army, U.S. Army Special Forces, Civil Affairs, and Psychological Operations teams. The first publication (I/ITSEC 2009) summarized findings from a quantitative study that investigated experiential learning in the multi-player, PC-based game module transitioned to PEO-STRI, DARWARS Ambush! NK (non-kinetic). The 2009 publication reported that participants of multi-role (Player and Reflective Observer/Evaluator) game-based training reported statistically significant learning and engagement. Additionally when the means of the two groups (Player and Reflective Observer/Evaluator) were compared, they were not statistically significantly different from each other. That is to say that both playing as well as observing/evaluating were engaging learning modalities. The Observer/Evaluator role was designed to provide an opportunity for real-time reflection and meta-cognitive learning during game play. Results indicated that this role was an engaging way to learn about communication, that participants learned something about cultural awareness, and that the skills they learned were helpful in problem solving and decision-making.

Problem Statement and Rationale

While the body of research regarding the study of tactical training game effectiveness is growing, empirical study of non-lethal, or non-kinetic, game effectiveness is lacking. With the current emphasis on culture and language training (Selmeski, 2007) more research is needed on soldier unscripted communication and intercultural negotiation in cross-cultural settings. Eventually pushing past tactical, kinetic, and scripted game effectiveness studies our community will better understand best practices for how we can make learning itself more effective with games.

The present paper is the second in a series that seeks to go beyond game effectiveness per se to understand how we can make game-based learning more effective, especially learning focused on training non-lethal, or non-kinetic engagement skills. Making learning more effective is an opportunity for out-of-the-box thinking about how to deliver and characterize a system of crucible experiences (Raybourn, 2007). For example, some games for training incorporate the use of multiple roles including roles for real-time reflection, evaluation, and feedback.

In two such cases, DARWARS Ambush! NK and Adaptive Thinking & Leadership, trainees are provided with opportunities to play multiple training roles designed to exercise unscripted intercultural communication, perspective taking, adaptive thinking, and metacognitive agility (Raybourn 2009a, b). The present paper presents our qualitative analyses of the unscripted transcripts from an empirical study investigating multi-role experiential learning (Raybourn, 2009a).

Research Objectives

The present paper seeks to understand salient communication behaviors of novice trainees of non-kinetic game-based missions by revisiting a quantitative study with further investigation such as stochastic player performance analysis using latent semantic analyses and graph visualizations. The qualitative analyses which are the subject of the present paper delve deeper into 78 trainees' in-game unscripted intercultural communication to further illuminate the results of a quantitative study reported at I/ITSEC 2009. We applied advanced analysis and visualization techniques to discover and interpret the underlying trends and patterns associated with communication behaviors that can someday potentially lead us to better predict how well trainees will perform in role-playing scenarios and whether game systems can respond to individual strengths and weaknesses as trainees learn communication strategies. The results are applicable to First-Person game-based learning systems designed to enhance trainee intercultural communication, interpersonal skills, and adaptive thinking.

MULTIPLE ROLES FOR EXPERIENTIAL LEARNING

The transcripts and evaluator ratings analyzed in the present study were generated during a 2009 empirical investigation of the use of multiple roles for experiential learning with games. The design of the multiple roles was inspired by Experiential Learning Theory combined modes for *grasping experience* via concrete experience and abstract conceptualization, and *transforming experience* via reflective observation and active experimentation (Kolb, 1984).

The experimental environment consisted of a multiplayer cordon & knock scenario in a fictitious environment. The DARWARS Ambush! NK platform is built on a commercial game (Operation Flashpoint, developed by Bohemia Interactive Studios) and includes roles for an instructor, soldiers, local nationals, observer/evaluators, and non-player-characters (NPC). Much like the Adaptive Thinking & Leadership training game (Raybourn et. al., 2005) role-players use headsets with microphones to communicate and interact with others during the game-based training. Reflective Observer/Evaluators provide real-time performance evaluations (Raybourn, 2009a).

In the experiment trainees in the role of the commander receive an Operations Order to bring a local national (LN) from the village back to the FOB for questioning. As a trainee the goal of this mission is to successfully conduct tactical questioning and negotiation in the intercultural setting. If successful, the trainee learns that the local national's cousin is an Imam who is cooperating with US Government. The questioning allows the trainee to practice aspects of intercultural communication such as cultural awareness, language, listening, cultural norms, and some nonverbal communication. The trainees' objective is to negotiate with the local national to return to the Forward Operating Base (FOB) for questioning willfully and voluntarily. Other tasks executed by the trainee include investigating a nearby marketplace where questionable equipment is for sale (e.g., weapons, night vision goggles) and communicating with merchants (non-player character text dialogue) who can provide additional information as to the whereabouts of the local national to be interviewed. Trainees interact with the interface similar to many commercial first-person perspective games including VBS 2 and others. The mission was designed to promote trainee learning via *grasping experience*.

Trainees in the role of the Reflective Observer/Evaluator were also briefed on the mission and asked to observe and evaluate the performance of the trainee in the role of the commander. Both roles were briefed on the evaluation criteria and each trainee had an opportunity to play both roles. The interface shown in Figure 1 illustrates how Reflective Observer/Evaluators track trainee performance. Evaluations are initiated by the instructor who selects the performance criteria from a competency drop-down list. The instructor sends requests for evaluation to all the Observer/Evaluators. Presently there are 10 general non-kinetic engagement competencies from which to choose including several items related to cultural awareness, leadership, communication, and adaptability. This role was designed to promote trainee learning via *transforming experience*.

Figure 1. Reflective Observer/Evaluator interface

SUMMARY OF 2009 FINDINGS

In the 2009 study we asked and answered the following research questions:

RQ1: Do participants, regardless of role (either player observation/evaluation role), report change with respect to their learning? Our results indicated that both roles (Player and Reflective Observer/Evaluator) reported statistically significant positive change in factors contributing to learning such as engagement and self-awareness. Additionally both groups reported that the skills learned during the training mission were helpful in solving problems and making decisions.

RQ2: Are there significant differences among groups participating in different roles in non-kinetic engagement training, especially when one role requires more active participation that the other? Our results indicated that while both groups reported statistically significant learning the trainees in roles of the commander also reported learning more about their own strengths and weaknesses. That is, when taken independently both groups reported statistically significant learning with players reporting learning about strengths and weaknesses, but when the means of the two groups were compared, they were not statistically significantly different. Groups differed when they should have (given the design of each role) with each group reporting statistically significant learning.

Detailed descriptions of the experimental environment, trainee multiple roles, and statistical analyses of findings are available (Raybourn, 2009).

QUALITATIVE STUDY METHOD

The present study asks the following research questions:

RQ3: What does the visualization of trainee performance data coming from unscripted game play transcription offer toward further understanding of the empirical results reported in the 2009 I/ITSEC paper?

RQ4: Can this approach be used for future games development of player/learner models and game adaptation algorithms?

Recall that the present paper seeks to understand salient communication behaviors of novice trainees of non-kinetic game-based missions by revisiting a quantitative study with further investigation such as stochastic player performance analysis using latent semantic analyses and graph visualizations. Additionally a long-term goal of this effort is to adapt a game automatically to a trainee, or create an individualized training vector. Our approach to this research is to capture total in-game task performance and attempt to derive or infer the individual's vector for the latter portion of the task performance from analyses of prior performance in the same session. This approach could allow us to adapt the game in real time— to change the latter segment of the game itself or to steer the trainee back to a pedagogically correct path. Our approach consists of constructing an algorithmic mapping from data to performance evaluation that models performance evaluations provided by a human rater/coder. We provide more detail on this procedure in the remaining sections.

Seventy-eight transcripts from a volunteer pool of 85 members of Sandia National Laboratories were analyzed for the present study. Seven transcripts were omitted from the study because they were incomplete, inaudible, etc. Sandia National Laboratories is an engineering laboratory located in Albuquerque, New Mexico. Most of the 78 participants in the present study were novices with little to no military experience, only 13% reported ever being or currently a member of the US Armed Forces. They ranged from ages 18 – 64. Only 12 females participated in the study. Sixty-two percent reported being European American. Thirty-three percent had master's degrees and 11% had doctoral degrees. Ninety-nine percent reported having no computer game-based training, although 64% had played single-player games, and 26% reported playing multi-player games.

Data Analysis Toolkit: Titan

The Titan Toolkit (Wylie & Baumes, 2009) addresses a growing trend that involves merging scientific visualization, which is physics-based and has a real-world representation, with information visualization, which is notionally abstract and has no well-defined representation. In the present paper, we used the algorithms developed in Titan to visualize and interpret military training in the form of game-play activity data combined with abstract qualitative meta-information (e.g., trainee performance or individual learning processes) in the game space. In doing so, we established a foundation on which to build a generalized utility for analyzing games and bridging from concrete game activities to qualitative meta-level performance and learner modeling.

Procedure

The initial step involved transcription of all the communication utterances in the game-based training mission. Two human rater/coders performed a content analysis rating the trainees on a scale of 1 to 5 as to how well they performed the task (see Table 1 for categories corresponding with 1 – 5 ratings). The videos of game play time-stamped events were also transcribed. The videos used in this study were from the Reflective Observer/Evaluator point of view so that comments from trainees in this role could be coded. The videos of game-play data were analyzed for type of speech (statement, question, answer, or comment.), category of speech (unscripted spoken or scripted text dialogue), and other information (i.e. vocal tone if any). In the category of "type of speech" a comment was help provided out of character by a research confederate or a trainee asking a question that did not pertain to the mission, such as "How do I use my gun?"

The videos were coded a second time by two different coders who were familiar with the training mission and had been trained to recognize core competencies of non-kinetic engagements in the game-based scenario. The inter-coder reliability score (Pearson's correlation coefficient, r =.79) was strong. The coders focused on salient events and noted the Reflective Observer/Evaluator ratings and comments. Examples of these ratings on a Likert-type scale (1= strongly disagree to 5= strongly agree) include items such as "Able to speak targeted phrases in foreign language" and "Effectively communicates with team." In some cases Reflective Observer/Evaluators also augmented rankings with open-ended statements. The human coders used transcripts, Reflective Observer/Evaluator ratings, comments, and trainee performance categories to rate the skill level of the trainees using a Likert-type scale (1 = unskilled player to 5 = skilled player). The final rating depended on trainees' ratings in the categories of Communication, Observation, Cultural Awareness, and Mission as well as trainees' overall effectiveness in gaining information, their use of communication strategies, and lastly the Reflective Observer/Evaluator's ratings. Table 1 illustrates the coding system used.

	Ratings				
	1	2	3	4	5
Communication	*Poor to little communication with Confederate *Poor to little communication with Host National *Doesn't Attempt to use Arabic greetings	*Poor communication with Confederate *Poor communication with Host National	*Okay communication with Host National *Needs prompting from Confederate *Doesn't utilize Confederate *Attempts Arabic phrases	*Good communication with Host National *Utilizes Confederate *Engages Confederate in mission *Attempts Arabic Phrases	*Uses Arabic Phrases *Good communication/utilizes confederate *Great communication with Host National
Cultural Awareness	*Poor to no Cultural Awareness	*Little Cultural awareness	*Attempts cultural awareness	*Demonstration of Cultural awareness	*Cultural Awareness more prominent across behaviors and communication
Observation	*Not aware of surroundings	*A little aware of surroundings	*Observant, but might not spot crowd or BOLO	*Might remember Fayez/BOLO *Aware of surroundings	*Concern for safety *Spots BOLO *Aware of crowd, binoculars or BOLO
Mission	*Narrow View of mission *Tries to arrest Host National *Forceful in getting Host National back to base *Forgets parts of the mission	*Tries to arrest Host National *Forceful		*Asks pointed questions *Concern for finishing mission non-kinetically	Asks pointed questions but is not forceful Concern for finishing mission non-kinetically

Table 1. Human Coding Process

Human Coding Results

There were four main categories evaluated that make up the aggregate 1-5 rankings that players received: Communication, Cultural Awareness, Observation, and Mission. Communication had subcategories such as communication with the Confederate and Communication with the Host National, attempting to use the Arabic Language. Observation included noticing the BOLO, and watching/noticing the crowd. Cultural awareness included remembering conversations with vendors and utilizing information learned in the market, expressing concern for civilians, etc. Mission included how the trainees viewed their mission and if they remembered mission objectives. Of the 78 trainee transcriptions coded, 7 trainees were rated a 1 (as characterized by the categories in Table 1 pertaining to 1 – 5), 16 were rated as 2, 17 were rated 3, 29 were rated 4, and 9 were rated 5. Transcripts were omitted if they were unable to be coded due to the video not picking up on player conversations or a video not completely recording a session.

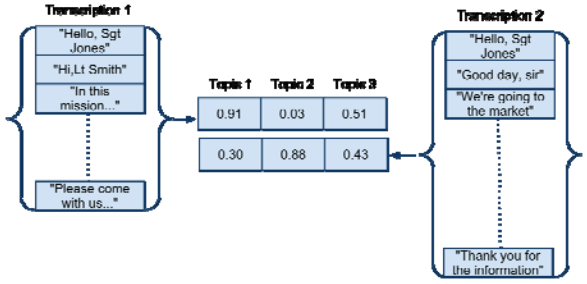

Figure 2. Transcription to Topic Space Diagram

Automated Coding Results

Following Landauer et al. (Landauer et al, 2009) to automatically score written text, we then used Titan to cluster the trainee's individual terms from the transcriptions collectively into topics via Latent Semantic Analysis (LSA) to produce vectors of weights over transcriptions in a fixed dimension topic space, (see Figure 2). For example, each trainee's transcription becomes a position in this topic space which enables an ordinary distance comparison between trainees' texts in the space. Using these topic-space vectors, we then applied standard machine learning techniques (Duda et al, 2001) to establish a model that maps trainees to scores. With the constructed model, we can predict scores for new trainees.

LSA (Landauer et al., 1998) works by taking a per-document term-frequency count and turning this into a set of topic weights. Each term is weighted for relevance to a topic, and similarly, each document is weighted for relevance to the same set of topics. This approach produces two sets of weights—one for mapping terms to topics and one for mapping documents to topics. We use the document weights to interpret similarity between players. Topics are chosen as a parameter to the system, in this case we settled on 9 topics by experimentation.

In order to account for the temporal aspects of the game data (similar to the approach found in Brants et al, 2002) each transcription is broken into three distinct sub-documents (see Figure 3). We chose these in such a way as to break up the final human-to-human negotiation that occurs within the game into three equally sized segments. For instance, if one player spent 15 minutes negotiating the segments would each by 5 minutes long. If another player negotiated for 20 minutes, that player's segments would be 6.6 minutes long. We decided to break up this final dialogue in this way, based on the human rating process where the final dialog was the most important factor in the overall score. Finally after running each sub-document through LSA we concatenated the vectors back into one longer vector of dimension 27.

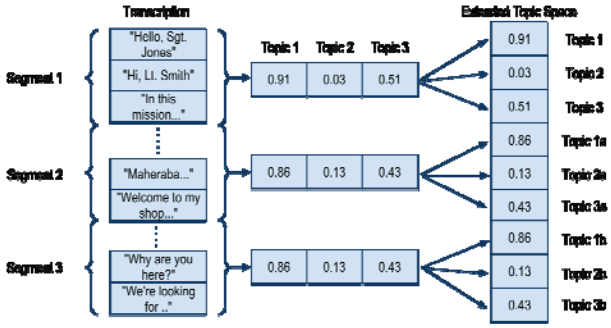

Figure 3 – Splitting Transcriptions Temporally

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
ah	shot	tent	shop	target	uh	uh	uh	uh
uh	hello	shot	inch	medical	bodies	confirmed	bodies	americans
nasir	safe	green	tv	neutralized	lieutenant	injured	village	leave
um	shop	crates	32	permission	return	rocket	tribe	tribe
sergeant	shooting	injured	greetings	lock	medical	care	leave	strange
okay	rpg	medical	models	clear	touch	bodies	terrorists	weapon
saiful	buy	night	flat	confirmed	happened	appears	members	cousin
hello	greetings	bodies	panel	force	throwing	permission	ambulance	night
talk	islam	tribe	ahead	uh	rocks	commander	crowd	village
help	injured	strange	electronics	commander	team	medical	green	understand

Figure 4. Top terms in each topic

Figure 4 shows the top ten weighted terms assigned by LSA to each of the nine topics. Note that, these topics correspond well with some of the categories the human judge used in the rating scheme, for instance Topic 5 matches well with "communication with confederate." In Figure 5 the association of document segments to topics, thus there are three vertices that belong to a document.

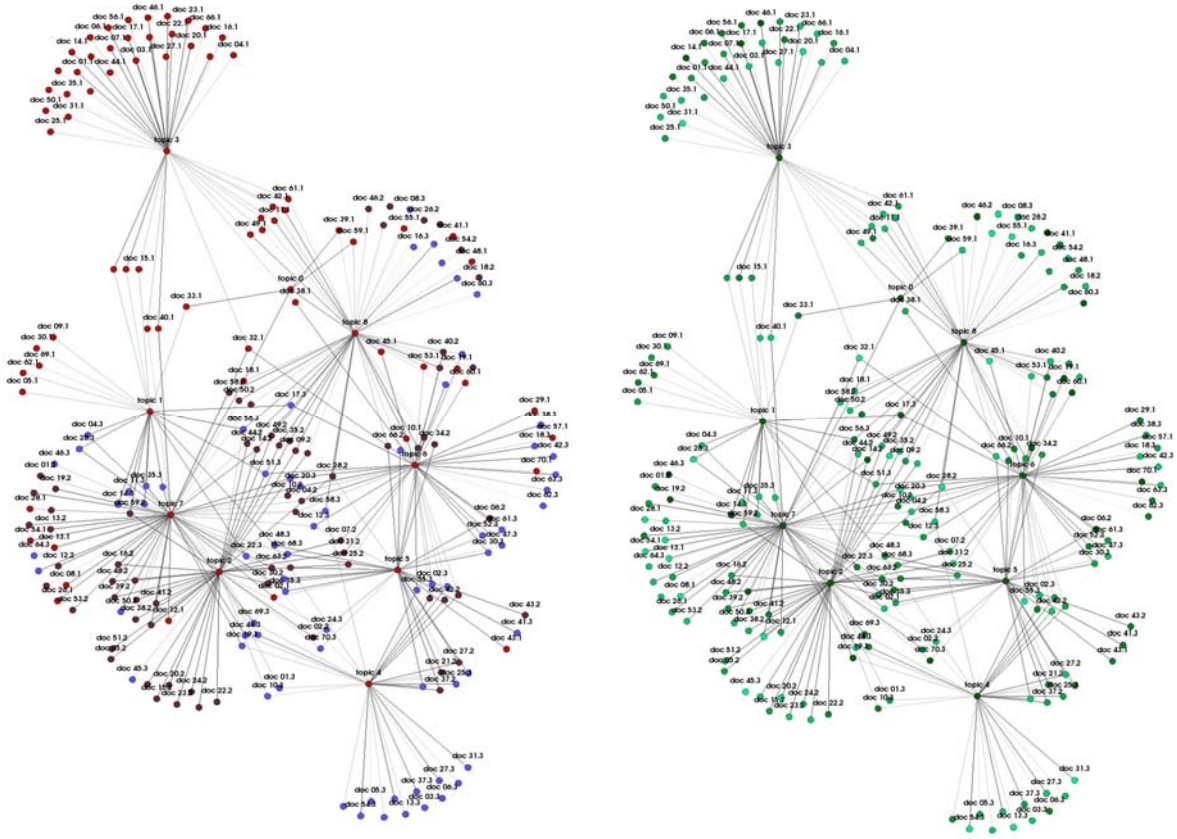

Figure 5. Top topics in document segments.
Left image vertices colored by segment. Right image vertices colored by human score.

Vertices are labeled with "doc x,y" where x is the document number and y is the segment number. In the left image we colored the vertices by the three segment numbers, from red to black to blue. In the right image, vertices are colored by the score (1 to 5, from dark green to bright green) given to the entire document. In Figure 5, there is no direct association between topics and rating. Although it is apparent from the graph on

the left that certain topics are important early, such as topic 3, and some late, such as topic 4, there is no clearly defined trend for all the topics strictly in terms of segments.

Figure 6 depicts the average topic weight across all documents at each segment in time. This shows again that topic 3 is strong in the early but not later segments. There are some patterns recognizable from Figures 4 and 5, but to complete the picture we elicited a more complicated mapping between topics and segments. To construct this mapping, we used a support vector machine (SVM) provided in the WEKA machine learning toolkit (Hall et al, 2009). This builds a classifier of our 27-dimensional topic-space vectors. We used the human rater's scores as train and test class labels.

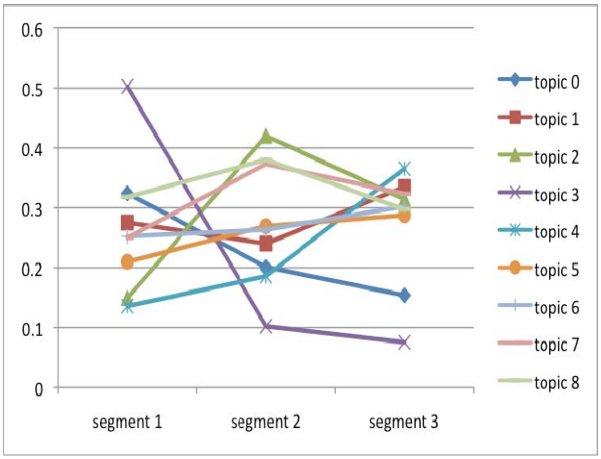

Figure 6. Average topic weight vs. temporal segment

RESULTS

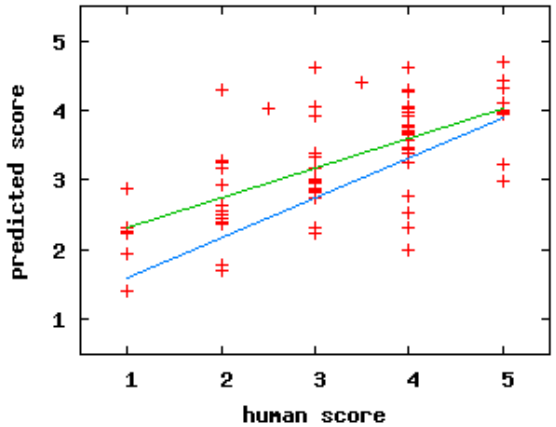

Figure 7. Correlation between LSA-SVM predicted scores and human scores

The model constructed by the SVM predicts the human rater's scores with an average Pearson's correlation coefficient of 0.63 over a 10-fold cross validation. The mean absolute error was 0.70. In other words, there was a disagreement of usually less than one point within the 1-5 Likert-type rating scheme used by the coders. The scatter view, in Figure 7, shows a comparison between the human rater's scores on the horizontal axis and the automatic prediction on the vertical. The upper green regression line through the data shows the regression between the predicted and actual scores. Inter-coder reliability with a second rater's scoring was 0.79 Pearson's correlation coefficient and its regression (without the corresponding points) is shown as the lower blue line in the scatter plot.

Despite not yet achieving the same level of correlation as that of the two humans, the automatic approach using LSA provides a strong baseline from which to improve. However, indications from our experiments with other methods of direct manipulation on the frequency data, e.g., reweighing and preclustering, or different algorithms (e.g., Latent Dirichlet Allocation, Neural Networks) increase the performance only minimally (or not at all). Thus, it is likely these results represent the highest performance available using a term-frequency approach in isolation.

CONCLUSIONS

Returning to the research questions of the present study: **RQ3—What does the visualization of trainee performance data coming from unscripted game play transcription offer toward further understanding of the empirical results reported in the 2009 I/ITSEC paper?** Now that we have evidence which supports using this approach to reliably corroborate human rater/coder ratings for unscripted communication in a non-kinetic engagement mission we believe visualizations can be used to further identify trainee strengths and weaknesses especially when against pre- and post-test self-report on learning. For instance, we could use the visualization to identify discrepancies between performance (as evidenced through game play behaviors, events, transcriptions, and Reflective Observer/Evaluator ratings) and self-report of perceptions of learning to better understand if any trainees perceive they learned even though their performance may have been rated poorly. We intend to explore this direction with a third contribution to this series of I/ITSEC papers.

With the goal of improving the accuracy of prediction in the present study, two avenues we intend to explore

are sentiment analysis (Lin and He, 2009) and automatic segmentation (Brants et al., 2002).

Sentiment analysis assigns a positive or negative score to terms using prior knowledge. This may lend itself to effectively weigh frequency counts and indicate whether the dialogue has a positive tone or a negative one. While LSA may do an effective job of clustering synonyms into topics, individual words within a topic may have dramatically different sentiment or tone. We expect such information would be crucial in a human rater's determination, conscious or not, of whether or not the dialog was successful.

By segmenting the transcription at three known points to incorporate a small amount of temporal information, we increased performance dramatically over applying LSA on the document as a whole. The approach used by Brants et al, (2002) is a fully automated method for segmenting a document into discrete regions. Using this in combination with sentiment analysis may give us a more complete picture of the conversation including when information is introduced to the discussion and the tone of the actors at various stages.

FUTURE RESEARCH

Finally regarding **RQ4—Can this approach be used for future games development of player/learner models and game adaptation algorithms?** As mentioned, the larger strategy guiding our incorporation of automatic scoring is to adapt a game's content automatically to a trainee's individual strengths and weaknesses. We are currently examining the effectiveness of temporal state space learning including the use of Dynamic Bayesian Networks (DBN) to create learner models. The automatic segmentation of transcriptions for LSA lends itself to construction of state space transitions, which are necessary as input to a DBN learning algorithm. We believe that this state-transition model is a useful general representation of activity in any game, exemplified by the more esoteric conversation space used in this paper, and thus as an approach, generally applicable.

The present paper sought to go beyond studies of game effectiveness to understand how experiential game-based learning systems can more effectively train non-kinetic engagement skills. To do so we introduced multiple learning roles that exercise different skills (reflection, perspective taking, adaptability, and intercultural communication).

We applied advanced analysis and visualization techniques to discover and interpret the underlying trends and patterns associated with communication behaviors that can someday potentially lead us to better predict how well trainees will perform in role-playing scenarios and whether game systems can respond to individual strengths and weaknesses as trainees learn communication strategies. The results are applicable to the assessment and design of First-Person game-based learning systems designed to enhance trainee intercultural communication, interpersonal skills, and adaptive thinking.

The ratings garnered from the Reflective Observer/Evaluator roles where used along with trainee transcripts, events, and coded behavior toward the assessment of trainee performance for the long-term goal of adapting training game content to account for individual strengths and weaknesses.

Finally our research supports the notion that experiential learning in game-based training systems may benefit from approaches designed to facilitate trainee mastery of modes for *grasping experience* via concrete experience and abstract conceptualization, and *transforming experience* via reflective observation and active experimentation (Kolb, 1984). As advanced in our 2009 I/ITSEC paper providing trainees with the opportunity to switch roles, and play from different perspectives we can engender the adaptive behaviors needed to excel in non-kinetic engagements and STTR operations.

ACKNOWLEDGEMENTS

*This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

REFERENCES

Beal, S. A. (2009). Exploring the Use of a Massive Multiplayer Game (MMPG) to Train Infantry Company Commanders. I/ITSEC 2009 Proceedings, Interservice/ Industry Training, Simulation and Education Conference Proceedings, November 30- December 04, Orlando, Florida, USA.

Beal, S. A. (2006). Lessons learned from evaluating training games for Infantry leaders. I/ITSEC 2006 Proceedings, Interservice/ Industry Training,

Simulation and Education Conference Proceedings, December 4-7, Orlando, Florida, USA.

Belanich, J., Orvis, K. L. Mullin, L. N. (2004). Training game design characteristics that promote instruction and motivation. I/ITSEC 2004 Proceedings, Interservice/ Industry Training, Simulation and Education Conference Proceedings, December 6-9, Orlando, Florida, USA.

Brants, T., Chen, F., & Tsochantaridis, I. (2002). Topic-Based Document Segmentation with Probabilistic Latent Semantic Analysis. In International Conference on Information and Knowledge Management (CIKM), McLean, VA, USA.

Brown, B. (2010). A Training Transfer Study of Simulation Games. Unpublished Master's Thesis, Naval Postgraduate School, Monterey, CA.

Duda, R.O., Hart, P.E., and Stork, D.G. Pattern Classification. John Wiley & Sons, 2001.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1.

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. New Jersey: Prentice-Hall.

Landauer, T. K., Foltz, P. W., Laham D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Landauer, T. K., Laham, D., & Foltz, P. (2003). Automatic Essay Assessment. Assessment in Education, Vol. 10, No. 3.

Lin, C. and He, Y. (2009). Joint Sentiment/Topic Model for Sentiment Analysis. CIKM '09: Proceedings of the 18t ACM conference on Information and Knowledge Management. 275-384.

Orvis, K., Horn, D. & Belanich, J. (2006). Videogame-Based training success: The impact of trainee characteristics - Year 2 (Technical Report 1188). U.S. Army Research Institute for the Behavioral and Social Sciences: Arlington, VA.

Raybourn, E. M., Deagle, E., Mendini, K., & Heneghan, J. (2005). Adaptive Thinking & Leadership simulation game training for Special Forces Officers. I/ITSEC 2005 Proceedings, Interservice/ Industry Training, Simulation and Education Conference Proceedings, November 28-December 1, Orlando, Florida, USA.

Raybourn, E.M. (2007). Applying simulation experience design methods to creating serious game-based adaptive training systems. *Interacting with Computers* 19, Elsevier. 207-14.

Raybourn, E.M. (2009a). Beyond Game Effectiveness Part I: An Empirical Study of Multi-role Experiential Learning. I/ITSEC 2009 Proceedings, Interservice/ Industry Training, Simulation and Education Conference Proceedings, November 30- December 04, Orlando, Florida, USA.

Raybourn, E.M. (2009b). Intercultural Competence Game that Fosters Metacognitive Agility and Reflection. In A.A. Okok and P. Zaphiris (Eds.). Online Communities, Lecture Notes in Computer Science (LNCS) 5621, 603-12. Springer-Verlag.

Rowan, P. A., & Brown, D. (2008). Games – Just how serious are they? I/ITSEC 2008 Proceedings, Interservice/ Industry Training, Simulation and Education Conference Proceedings, December 1-4, Orlando, Florida, USA.

Selmeski, B. (2007). Military cross-cultural competence: Core concepts and individual development. Armed Forces & Society Occasional Paper Series, (1) AFCLC Contract Report 2007-01. Canada: Royal Military College of Canada Centre for Security.

Wylie B., & Baumes, J. (2009). A Unified Toolkit for Information and Scientific Visualization. In Proceedings of Visualization and Data Analysis (VDA). SPIE.