Metrics Assessment toward a Training Effectiveness Evaluation of Augmented Virtuality for Call for Fire Training: Insights from a Novice Population

Julie N. Salcedo
Design Interactive, Inc.
Orlando, FL
julie.salcedo@designinteractive.net

Stephen R. Serge
Institute for Simulation & Training
University of Central Florida
Orlando, FL
sserge@ist.ucf.edu

Stephanie J. Lackey
Design Interactive, Inc.
Orlando, FL
stephanie.lackey@designinteractive.net

Roberto Champney
Design Interactive, Inc.
Orlando, FL
roberto@designinteractive.net

Gino Fragomeni
Army Research Lab
Simulation and Training Technology Center
Orlando, FL
gino.f.fragomeni.civ@mail.mil

Jonathan Hurter
Institute for Simulation & Training
University of Central Florida
Orlando, FL
jhurter@ist.ucf.edu

ABSTRACT

Call for Fire (CFF) is a highly complex and dynamic task to train. Existing CFF training systems offer immersive training experiences, yet high set-up and implementation costs and limited system portability inhibit fulfillment of throughput requirements. Augmented Virtuality (AV) may be a viable solution to reduce costs associated with CFF Simulation-Based Training, improve system portability, increase throughput, and enhance the immersive experience. AV involves the blending of live and virtual training elements to create a highly immersive experience with greater task fidelity. This experiment represents an initial metrics and experimental protocol assessment in a series of training effectiveness evaluation experiments investigating the performance and learner perception tradeoffs of AV technologies applied to the CFF task domain. Results reveal trends toward increased learner self-efficacy, positive perceptions of system fidelity and usability, and high ratings for immersion, engagement, and presence. These findings confirm the validity of the selected performance metrics and subjective measures for the assessment of AV technologies for CFF training and also inform the empirical recommendations to improve the quality of follow-on training effectiveness evaluations.

ABOUT THE AUTHORS

Julie N. Salcedo, Ph.D. earned her Master's and Ph.D. in Modeling and Simulation and a Certificate in Instructional Design for Simulations from the University of Central Florida (UCF). Dr. Salcedo is a Research Associate at Design Interactive, Inc. A former public school educator for the state of Florida, Dr. Salcedo leverages her education, instruction, and assessment background to investigate instructional strategies and design features for simulation-based training systems and evaluate the impact on training effectiveness. The focus of her research efforts involve assessing the training effectiveness of emerging simulation-based training platforms, including Augmented Reality (AR), Augmented Virtuality (AV), and Virtual Worlds (VW), applied to the Intelligence, Surveillance, Reconnaissance (ISR), human terrain analysis, team and squad maneuver, and Call for Fire (CFF) military task domains.

Stephen R. Serge, Ph.D. is a researcher with UCF IST specializing in training effectiveness, instructional design, human-technology interaction, and usability. Dr. Serge has been directly involved with game and SBT research and design for over 5 years and has assisted in the development of intuitive user interfaces on numerous projects. He received his M.A. and Ph.D. in Human Factors Psychology from UCF in 2012 & 2014, respectively. Presently, his research efforts focus on assessing the usability and utility of current and emerging Head-Mounted Display (HMD) technologies for immersive SBT.

Stephanie J. Lackey, Ph.D. earned her Master's and Ph.D. degrees in Industrial Engineering and Management Systems with a specialization in Simulation, Modeling, and Analysis at UCF. Dr. Lackey has over 15 years of experience conducting innovative and high-risk research aimed reducing the risk of the downstream acquisition enterprise. She serves as the Director of the Federal Solutions Division at Design Interactive, Inc. and leverages her experience in advanced predictive modeling and simulation-based training in order to develop methods and products that optimize human performance.

Roberto Champney, Ph.D. is Chief Scientist - Engineering at Design Interactive, Inc. His work focuses on field analysis, design and evaluation of interactive and training systems, and emotion in design. Dr. Champney holds a Ph.D. degree in Industrial Engineering and Management Systems from UCF.

Gino Fragomeni serves as a Science and Technology Manager for Dismounted Soldier Technologies at the U.S. Army Research Laboratory, Simulation and Training Technology Center. He currently works with the Ground Simulation Environments Branch conducting research and development in the area of dismounted Soldier training and simulation. Mr. Fragomeni's research interests include artificial intelligence and immersive environments centric to dismounted training applications. Currently, he is conducting a training effectiveness evaluation to assess the efficacy of AV technology for highly immersive Joint Forward Observer (JFO) training in the CFF task domain. Mr. Fragomeni is a highly qualified science and technology manager as well as being a Sergeant Major with 3rd Battalion 20th Special Forces Group with over 31 years of military experience. He earned a Master of Science from UCF and has specialized training in Systems Engineering.

Jonathan Hurter earned a Bachelor's degree in Digital Arts (Sound Track) from Stetson University in 2012, and is currently pursuing a Master's degree at UCF in Modeling and Simulation and a Certificate in Instructional Design for Simulations. Mr. Hurter's recent interests include instructional design, serious games, and new media for learning. He was awarded a Provost's Graduate Merit Fellowship in 2014, and was a panelist for online learning at the UCF Center for Distributed Learning. As a current Graduate Research Assistant for the ACTIVE Lab, Mr. Hurter supports current SBT research focused on AV and VW technologies.

Metrics Assessment toward a Training Effectiveness Evaluation of Augmented Virtuality for Call for Fire Training: Insights from a Novice Population

Julie N. Salcedo
Design Interactive, Inc.
Orlando, FL
julie.salcedo@designinteractive.net

Stephen R. Serge
Institute for Simulation & Training
University of Central Florida
Orlando, FL
sserge@ist.ucf.edu

Stephanie J. Lackey
Design Interactive, Inc.
Orlando, FL
stephanie.lackey@designinteractive.net

Roberto Champney
Design Interactive, Inc.
Orlando, FL
roberto@designinteractive.net

Gino Fragomeni
Army Research Lab
Simulation and Training Technology Center
Orlando, FL
gino.f.fragomeni.civ@mail.mil

Jonathan Hurter
Institute for Simulation & Training
University of Central Florida
Orlando, FL
jhurter@ist.ucf.edu

INTRODUCTION

Call for Fire

A Call for Fire (CFF) is a transmitted request for indirect fires to suppress, neutralize, mark, or destroy enemy targets by a Joint Forward Observer (JFO) in the field (UNC Charlotte ROTC, 2010; U.S. Army, 1991). JFOs are Soldiers with specialized knowledge and skills that allow them to effectively request fires on a threat or target and serve as a critical link between dismounted Soldiers and fires support units (Stensrud, Fragomeni, & Garrity, 2013). A CFF mission typically involves a JFO located at an observation post in close proximity to the mission or target area to allow for a clear and continuous view of the objective. From this position, the JFO acquires target information, communicates that information to a Fires Direction Center (FDC), who verifies and relays the information to the fires support team, who then fires munitions on the location provided (U.S. Army, 1991). During indirect fire requests, the JFO is responsible for providing the position and target description to the FDC, as well as spotting impacts and adjusting subsequent firing attempts, if necessary (Stensrud, Fragomeni, & Garrity, 2013; U.S. Army, 1991).

The CFF task consists of several identification, communicative, and assessment subtasks, each requiring a high level of efficiency and accuracy. The task itself requires, at minimum, three transmissions consisting of six total elements required for proper execution of a CFF mission (U.S. Army, 1991). Initial communication from the JFO to the FDC consists of JFO identification and a warning order. These allow the FDC to prepare for a specific type of fires request from the JFO (i.e., location of JFO, method of geo-locating targets, etc.). The next communication requires the JFO to send target location, determined using one of three location finding methods (i.e., grid, polar, shift), and confirm coordinates with the FDC. The final element in a CFF consists of the target description, the requested method of engagement, and the method of fire and control (e.g., “at my command,” “time on target”). After the third transmission, the FDC forwards approved firing commands to the firing unit, who is responsible for firing on the requested target. After completion of the CFF request, the JFO monitors and reports effects on targets, requests adjustments for additional fires, or ends the mission and reports a battle damage assessment.

Training Challenges

Training for JFO certification includes multiple phases in classroom, live, and/or simulated exercises. Initial training involves learning the capabilities and weapon systems of the various attack vehicles, the function and procedures needed for effective CFF tool utilization (e.g., maps, compasses, etc.), and the proper terms and methods of radio communication during a CFF mission. Each of these must be learned and executed to proficiency before a JFO receives certification. In addition, sustainment training is necessary for certification maintenance. Unfortunately, access to adequate rehearsal is sometimes limited due to extraneous factors, such as funding availability, limited resources, or time constraints (Ryan, 2011). Decreased availability of live-fire or inadequate simulated rehearsal potentially leads to an increased risk of current JFOs falling out of certification, which results in fewer deployment-ready JFOs than expected. Long durations of nonuse of trained knowledge and skills presents a risk of skill decay particularly related to proficiency and performance (Wang, Day, Kowollik, Schuelke, & Hughes, 2013). The three major predictive task-related indicators of skill decay include: discrete (e.g., CFF, rifle assembly) versus continuous (e.g., driving, running), high cognitive demand, and long retention intervals (i.e., time between training and testing/application; Wang, Day, Kowollik, Schuelke, & Hughes, 2013). Often, the most effective methods for maintaining high proficiency, and minimizing skill decay, are highly effective initial training and the ability to rehearse skills frequently (Ryan, 2011; Wang, Day, Kowollik, Schuelke, & Hughes, 2013). The utilization of cost effective and easy-access simulated CFF training is being explored to address both of these factors.

Current CFF Training Systems

A current standard for U.S. Army CFF training utilizes the Call For Fire Trainer-II (CFFT II), an individual or collective Simulation-Based Training system that provides a simulated battlefield environment for training JFOs at the institutional and unit level (U.S. Army PEO STRI, 2012). The system is designed to be transportable and provide advanced distributed learning, simulated military equipment, a Virtual environment (VE), and computer-generated forces using One Semi-Automated Forces (OneSAF). The VE is projected on a large projector screen. Trainees interact with the VE by utilizing the simulated military equipment (e.g., laser designator rangefinder, night vision goggles, binoculars), a map, binoculars, and a student control computer to execute the scenario mission at the student station. An instructor facilitates training and plays the role of the FDC through an instructor station.

During scenario execution, trainees are often required to shift attention between virtual and real worlds throughout the training process, leading to incorrect usage of equipment and complaints from trainees that the simulation lacks an expected level of realism (Fragomeni, Lackey, Champney, Salcedo, & Serge, 2015). This presents a challenge in maintaining immersiveness throughout the mission. The benefits of increasing the immersive characteristics of CFF training simulations include the potential to improve trainee engagement and feelings of presence during training experiences (Van der Land, Schouten, Feldberg, Van den Hooff, & Huysman, 2013; Gamito, Morais, Oliveira, Gamito, & Anastácio, 2006), which has shown to also help increase training effectiveness (Jackson & McNamara, 2013). In order to accomplish this goal, it is necessary to integrate and examine emerging technologies for CFF training. Augmented Virtuality (AV) may be a viable solution to increase realism, engagement, and training effectiveness of CFF training systems.

Augmented Virtuality

AV is a type of Mixed Reality (MR), the latter falling under the umbrella of Virtual Reality (VR). The reality-virtuality continuum (Milgram & Kishino, 1994) places Augmented Reality (AR) and AV as two extremes of MR, with AR proximate to physical environments, and AV proximate to virtual environments. In AR, real or unmodelled environments are superimposed with virtual assets, such as synthetic objects, to alter a user experience. The virtual objects occlude the real elements. In contrast, AV is a reverse structuring of AR, where a predominately virtual environment introduces tangible elements, such as real objects. An example of AV includes the use of telepresence, where live video feeds are projected inside a Virtual World (Hughes & Stapleton, 2005; Regenbrecht et al., 2004).

An avenue for increasing immersiveness and integrating AV technology is via the incorporation of a see-through head-mounted display (HMD), which allows trainees to view a simulated environment while also maintaining the ability to see and use physical tools during training. HMDs are a means of increasing immersive characteristics of SBT (Moss & Muth, 2011). Since the HMD renders the virtual environment in response to head movements, users can look around themselves to obtain environment awareness in a natural fashion. In addition, the see-through component of the HMD allows for better integration of the JFO equipment (e.g., lensatic compass, binoculars, terrain maps), increasing realism and helping to maintain domain specific procedures. Each object is outfitted to merge with the virtual environment, actions performed with the tools have a real-world correlate, and the boundaries between the real and virtual elements deteriorate. For the CFF task domain, this allows for natural tool use, especially for tasks of self-location, target location, and adjustments by an observer. Detecting the objects' voids in a scenegraph allow the physical objects to be seamlessly embedded within the synthetic environment. AV also leverages the adaptiveness of digital virtual environments to supply a wide range of terrain characteristics, and target locations.

Purpose of the Experiment

The purpose of this experiment was to conduct an initial pilot assessment of performance metrics and subjective measures utilized toward the evaluation of AV technologies for CFF training. This pilot experiment was executed with a representative novice population, whose participation also assisted in the assessment of the experimental protocol. Results of this pilot experiment contributed to validation of assessment variables and recommendations to improve the experimental design of follow-on experimentation for this AV training effectiveness evaluation effort.

METHOD

Participants

A total of five cadets, four male and one female, in the Reserve Officers' Training Corps (ROTC) from a large southeastern university, with an average age of 22.6 years (*SD* = 2.51), participated in the evaluation. These participants were considered a representative sample for this experiment because of their general inexperience with the CFF task and their domain relevant backgrounds (i.e., military). Participation was restricted to U.S. citizens between the ages of 18-40 with normal, or correct-to-normal, vision, and full color vision.

Experimental Materials and Design

This experiment utilized the Call for Fire Trainer-Augmented Virtuality (CFFT-AV) prototype as the SBT testbed. The CFFT-AV system consists of a 7'x7'x7' frame equipped with an overhead positional tracking system, simulated military equipment (e.g., compass, binoculars, etc.), a partially occluded HMD, and a table with a terrain map and writing instruments. Participants wore the HMD to view a virtual environment depicting a JFO's observation post. Participants were able to use the additional equipment and instruments as necessary by looking at or picking up and holding the appropriate item within the field of view of the HMD. Participants stood within the CFFT-AV frame in front of the table during each scenario.

Table 1. Basic procedures to execute a Call for Fire mission.

Initial Procedures	1. Determine grid coordinates/location of observation post.
	2. Correctly detect and ID target as an enemy.
First Radio Transmission	3. Conduct a radio check with FDC.
	4. Send observer ID and location to FDC.
	5. Transmit the Warning Order (WO) to the FDC and wait for confirmation.
	- Contains mission type (e.g., Adjust Fire) and target location method (e.g., Polar).
	6. Await conformation from FDC.
Second Radio Transmission	7. Determine and transmit the target location information.
	- Determine direction, estimate distance, and determine vertical shift of target.
	8. Await confirmation from FDC.
Third Radio Transmission	9. Observer describes target in detail.
	10. Determine the method of fire and control.
	11. Receive confirmation of information from FDC.
	12. Receive "Shot, over" warning from FDC. Return "Shot, out" to FDC.
Observation, Adjust, and Report	13. Observe impact.
	14. Determine proper adjustments of fire. Transmit adjustment information to FDC.
	- Repeat steps 7-14 as necessary, based on impact of previous round.
	15. Determine effects on target and assess collateral damage. Report to FDC.

Participants performed in multiple virtual scenarios that required the accurate and effective execution of the CFF task procedures. A summary list of these procedures is located in Table 1. The testing variable for the current experiment was the use of the CFFT-AV for training the CFF task. All participants received identical simulation experiences.

Dependent Variables

Dependent variables included performance metrics and subjective measures. Performance scores consisted of accurate task execution, focusing on location deviation and completion time scores. Metrics for successful completion were provided by SMEs and qualified CFF training personnel. Each participant was scored on six performance metrics. Each was scored on a "Go/No-Go" mark for readiness (i.e., pass/fail). In order to obtain a "Go," participants must have completed the metric correctly and within the allotted timeframe; a "No-Go" was given if any step during the CFF task is incorrect or incomplete.

A number of subjective ratings were also collected. The Self-Efficacy Measure was a 29-item questionnaire that assessed the current confidence level of an individual participant at performing the CFF task, rated on a self-report scale between 0-100 (i.e., not confident to extremely confident, respectively). Self-efficacy responses were collected before and after exposure to the CFFT-AV. This measure was used to gauge changes in subjective self-efficacy based on simulation exposure.

A 20-item Learner Reactions Questionnaire allowed individuals to rate their perceived learning benefits from using the AV system to learn the CFF task on a 7-point (i.e., "Strongly Disagree" to "Strongly Agree"). The Learner Reactions Questionnaire was developed specifically for assessment of training systems applied to the CFF task domain. The items of this measure align with the learner reactions factor of training effectiveness evaluation as outlined by Kirkpatrick and Kirkpatrick (2006).

Simulator fidelity ratings were collected on an 11-item survey and on a scale from 1-5. The measure was used to obtain ratings on the perceived realism of the task, environment, and tools used in the CFFT-AV system. Usability was assessed at a system level using the usability subscales of the Technology Acceptance Measure (TAM; Zhang, Li, & Sun, 2006). Additionally, usability was assessed at the task level with the Task Execution Questionnaire, which was developed for usability assessment of CFF training systems. Together, the TAM and Task Execution Questionnaires examined how well interaction with the various facets of the CFFT-AV system and task were intuitive or pleasurable for the participants.

Levels of immersion, engagement, and presence were also collected on a series of three questionnaires with various scales (Jennett et al., 2008; Charlton & Danforth, 2005; Witmer, Jerome, & Singer, 2005). Higher scores on an individual measure indicated higher levels of the respective response regarding interactions with the simulation environment and tools.

Finally, the Simulator Sickness Questionnaire (SSQ) was used to examine if any participants experienced symptoms of simulator sickness throughout the experimental session. The SSQ contains 16 items that enable participants to indicate the severity of symptoms related to disorientation, nausea, and oculomotor disturbance experienced at the time (Kennedy, Lane, Berbaum, & Lilienthal, 1993).

Procedures

Individual participation took place over the course of a single day. Participants first reviewed consent information and were able to ask any questions regarding their participation. After providing voluntary consent, participants received classroom-based training using traditional CFF training content and methods. Training was facilitated by a SME with both training and field experience in the CFF task domain. The training session lasted approximately 3 hours. Next, participants completed a demographics questionnaire (e.g., age, military experience, familiarity with task-relevant tools, etc.), pre-task Self-Efficacy Measure, and a baseline SSQ.

Upon completing the initial questionnaires, participants learned how to operate and interact with the various tools and environment within the CFFT-AV simulator. After familiarization with the simulation equipment, participants performed in a series of two simulator sessions in the CFFT-AV system designed to test performance in accordance with the standard CFF evaluation criteria. The first session consisted of two training scenarios and permitted the SME to provide domain feedback. The second session consisted of one evaluation scenario and limited SME feedback. The SME also acted the role of the FDC to provide the necessary radio responses to support mission execution. The experimenter noted participant communications and physical actions throughout the simulated CFF scenarios. In addition, audio and video recording equipment was used to further document participant communication and actions during scenario execution in order to aid in accurate scoring of procedural performance. A midpoint SSQ was administered between the first and second simulator sessions.

Once a participant completed all of the CFF scenarios, they answered set of final questionnaires that included the post-task Self-Efficacy Measure, post SSQ, Learner Reactions Questionnaire, simulator fidelity survey, TAM, Task Execution Questionnaire, and the immersion, engagement, and presence questionnaires. After completion of the final questionnaires, participants received a debriefing, had an opportunity to ask any questions, and were dismissed.

RESULTS

Participants rated their familiarity with the specific tools used for the CFF task on a scale between 0-4. Two participants reported having prior CFF training, but none had performed a live CFF or received certification as a JFO. Demographic data revealed a moderate level of CFF-related tool familiarity (M = 2.45, SD = .758; maximum '4').

Simulated CFF Task Performance

Results from the performance scenario conducted in the CFFT-AV simulator revealed that participants were able to pass a number of the "Go/No-Go" criteria for performing a polar CFF mission. Examining individual task metrics revealed that, collectively, participants passed 80% of the self-location, initial estimation of target direction and distance, and the threshold for number of called corrections required for a "Go" classification. However, no participants were able to identify their location or identify the target within the allowable timeframe of 2-minutes for a "Go" classification, leading to an overall combined pass rate of 53% (i.e., location accuracy, distance estimation, and time). Table 2 shows the pass/fail performance for each metric.

Table 2. "Go" and "No-Go scores for performance metrics.

Participant	Self-Location Deviation (meters)	Target Direction Deviation (mils)	Target Distance Deviation (meters)	Number of Corrections	Time to Locate Self (sec)	Time to Locate Target (sec)
1	100	30	0	3	173*	234*
2	1100*	20	100	4	826*	180*
3	100	30	100	3	432*	556*
4	1000*	80*	0	2	429*	255*
5	100	80*	0	2	230*	342*
Objective	≤ 100	≤ 50	≤ 250	≤ 4	≤ 120	≤ 120

Note. * indicates "No-Go."

Subjective Measures and Perceptions

Subjective self-efficacy for the CFF task was obtained through survey responses collected between the classroom training and simulator session and after completion of all scenarios in the CFFT-AV system. Participants rated their ability to perform specific tasks required for a CFF mission on a scale of 0% – 100%. Results from a pair-samples t-test indicated a significant increase in self-efficacy ratings taken before (M = 66.62, SD = 16.16) and after (M = 83.59, SD = 4.56) the simulator sessions were completed, t (4) = 3.19, p = .033, d = 1.43. Overall, there was a 17% increase in self-efficacy for the CFF task after the simulator experience.

After completing the entire training session, participants were asked to complete the Learner Reactions Questionnaire. Ratings on items ranged from '1' (i.e., strongly disagree) to '7' (i.e., strongly agree). Analysis revealed that participants considered the CFFT-AV system moderately positive in terms of learning (M = 5.91, SD = 1.03). A one-sample t-test revealed that this was significantly higher than a response at or below the "Neutral" rating of '4' on the rating scale, t (4) = 4.19, SD = 1.03; d = 1.85.

In addition, participants also provided some insight regarding their perception of the system fidelity. Each participant rated fidelity on 11 simulator and task specific qualities incorporated into the CFFT-AV system. Ratings were provided on a scale of '1' (i.e., strongly disagree) to '5' (strongly agree). Descriptive results indicated general agreement that the CFFT-AV system provided adequate and realistic levels of fidelity related to the CFF task, with an overall average rating of 4.34 (SD = .29). Examining individual item ratings revealed that only "Audio Realism" was rated below '4' (M = 3.9, SD = .20), which is likely due to the limited audio capability at the time of the experiment. Overall, the ratings for system fidelity were significantly higher than a 'Neutral' rating of '3' on the rating scale, t (4) = 8.56, p = .001, d = 3.83. Figure 1 depicts the results of the fidelity ratings.

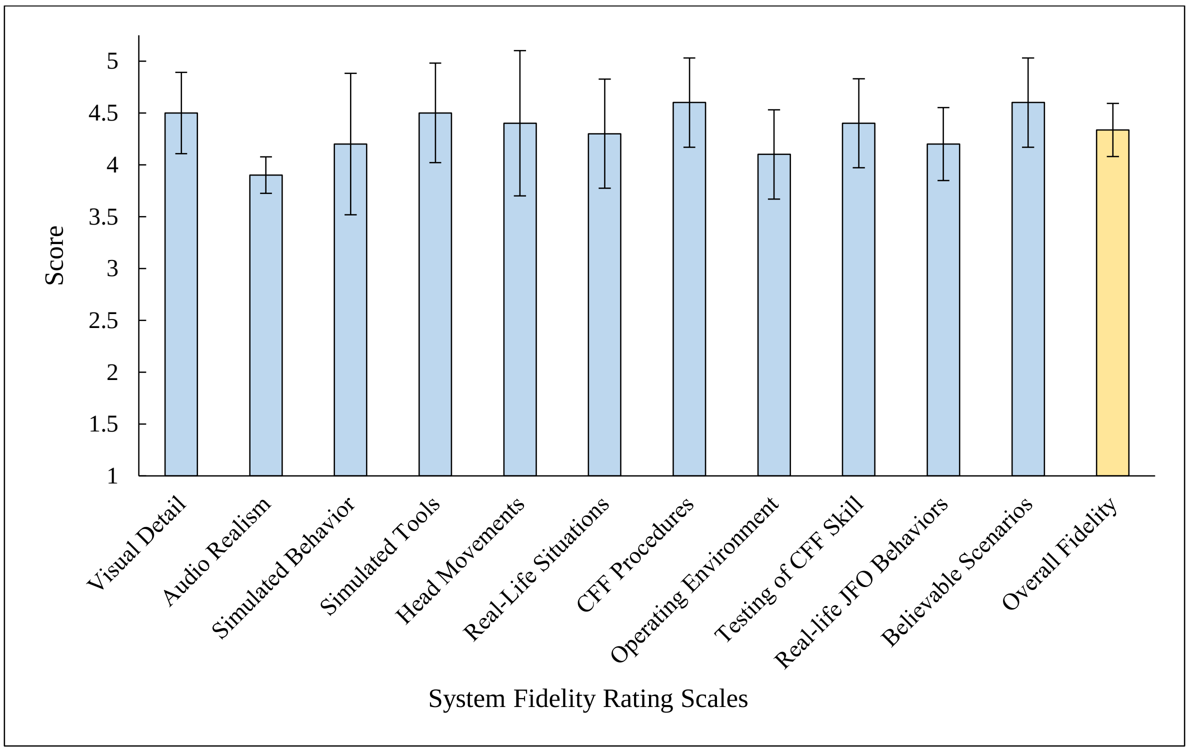

Figure 1. System fidelity ratings for the CFFT-AV.

Usability was measured using two questionnaires. The first consisted of the usability scales of the TAM, an 8-item survey measured on a scale from 1-7 (i.e., strongly disagree – strongly agree) that assessed overall system usability. Participants had a mean rating of 5.83 (SD = 1.06). This score was significantly above the "neutral" rating of '4,' trending towards the positive side of the scale, t(4) = 3.85, p = .018, d = 1.72). The Task Execution Questionnaire measured task specific usability factors. This measure was positively correlated with the TAM (r = .92, p = .03) and also indicated a significantly positive rating (M = 4.37, SD = .428) when compared to its "neutral" scale score of '3,' t (4) = 7.17, p = .002, d = 3.21.

Next, immersion, engagement, and presence were examined. Results for immersion indicated that, on a scale ranging from 0-32, participants average score was 26.2 (SD = 2.76). This rating was significantly higher than a scale midpoint score of 16, t (4) = 8.22, p = .001, d = 3.68. Similarly, engagement results indicated positive levels of participant engagement while using the CFFT-AV (*Median* = 34.00, *Range* = 3; max score of 35). These scores were significantly greater than a median "neutral" rating of 17.5 ($Z = 15.00$, $p = .039$). The items on the presence questionnaire were rated on a scale from 1-7, with a rating of '4' serving as the "neutral" point. Results showed a mean of 4.9 ($SD = 0.65$), which was significantly higher than "neutral," $t\ (4) = 3.12$, $p = .036$, $d = 1.39$.

Feelings of simulator sickness were also recorded before simulator exposure (i.e., baseline), between the training and evaluation simulator sessions (i.e., midpoint), and after all exposures to the simulator (i.e., post). No significant differences were observed from baseline ($M = 6.73$, $SD = 6.69$) through the midpoint ($M = 13.46$, $SD = 12.57$) and post ($M = 4.49$, $SD = 6.69$) measurements, $F\ (2,8) = 2.10$, $p > .05$. Figure 2 presents data from the individual participants and mean scores for the three SSQ administrations. The highest possible weighted score for the SSQ was 246.84. Overall, the level of simulator sickness was low, with minor variations between each administration.

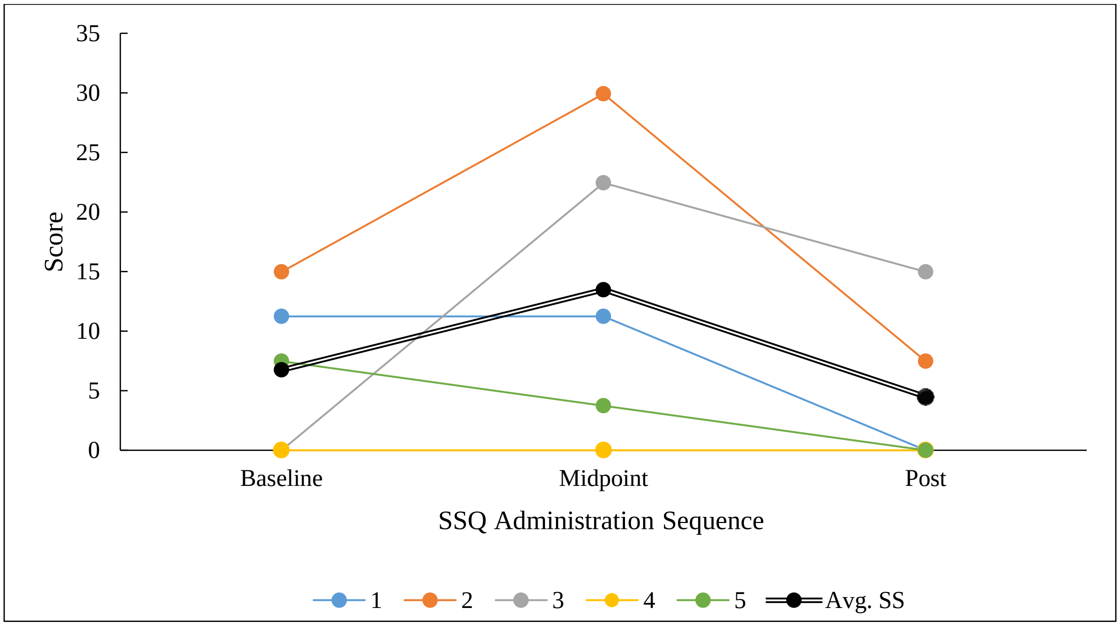

Figure 2. Individual and average responses for the baseline, midpoint, and post SSQ administrations.

DISCUSSION

Given the limited sample size for this experiment, it is difficult to diagnose whether level of experience or the AV technologies directly impacted performance results, particularly timed metrics (i.e., Time to Locate Self and Time to ID Target). The participants represented CFF task novices, therefore, it is possible that "No-Go" performance scores were simply a consequence of inexperience with the task. However, there may be an element of the simulation that impacted performance. Follow on data collection using participants from a broader range of CFF experience levels may reveal greater variability in performance results, which should assist in delineating the impact of task experience and AV technology factors (Lackey, Champney, Fragomeni, Salcedo, & Serge, 2015). Additionally, follow on investigations should explore implementing performance assessment methods beyond the binary "Go/No-Go" scoring. In empirical training effectiveness research, binary scoring may be limited in its ability to properly assess levels of task understanding and skill development (Lackey, Salcedo, Matthews, & Maxwell, 2014). Perhaps, scaled or percentage scoring methods will provide more descriptive performance scores, which may assist in deriving a more definitive interpretation of performance outcomes.

Self-efficacy, learner reactions, system fidelity, and usability results indicate an overall positive response to practice and training with the CFFT-AV. Self-efficacy for the CFF task increased after training sessions in the CFFT-AV simulator, which may be due to the practice afforded by SBT, in general, or to the implementation of AV technologies, specifically. Positive responses from relative novices on the Learner Reactions Questionnaire suggest the AV training system was a welcome training medium. System fidelity ratings trended toward the positive end of the rating scales indicating that participants likely accepted the level of realism provided during CFFT-AV scenarios. Moving forward, the collection of self-efficacy, learner reactions, and system fidelity perceptions from a larger sample size may contribute to the prioritization of training objectives and fidelity design recommendations for emerging AV training systems applied to the CFF domain.

Immersion, engagement, and presence results also trended toward higher, positive ratings, however, it is difficult to distinguish whether these variables were affected by the AV technology set-up or by the amount of SME involvement during the simulator sessions. Prior SBT research has shown that for many training tasks the involvement of a skilled instructor is often critical to the successful implementation and effectiveness of a SBT system (Crovella & Lipsky, 1997; Oser, Gualtieri, Cannon-Bowers, & Salas, 1999; Wray, Laird, Nuxoll, Stokes, & Kerfoot, 2005; Lackey, Salcedo, Matthews, & Maxwell, 2014). Given the novice population of this experiment, it was necessary to include a qualified SME to provide domain feedback to mitigate any risk of negative training due to participants' limited task experience. Evidence from the SBT literature indicates that SBT paired with skilled instruction has shown to improve operational performance speed by 84% and reduce errors by 72% (Haque & Srinivasan, 2006). Therefore, in order to better diagnose impact of AV on immersion, engagement, and presence, follow on experimentation should utilize a more experienced sample population where the risk of negative training is negligible so that the experimental design can significantly reduce any bias of SME involvement by standardizing the domain feedback across participants.

The relatively low simulator sickness scores indicate that the CFFT-AV did not induce detrimental or excessive levels of nausea, disorientation, or oculomotor interference. However, the fluctuations in simulator sickness between SSQ administrations may be an indicator of the impact of the experimental design. The midpoint SSQ was administered after the first simulator session which consisted of two training scenarios, while the post SSQ was completed after the second simulator session with only one evaluation scenario. During the initial session, the SME played the role of the FDC and was also permitted to interject and assist participants with task related feedback. This protocol doubled the anticipated time of 12 minutes per training scenario to an average of 24 minutes per scenario across participants (i.e., initial session total of 48 minutes). Prolonged exposure to simulated environments has shown to increase the instance of simulator sickness (Classen, Bewernitz, & Shechtman, 2011). Further, the slight peak in simulator sickness at the midpoint measurement is consistent with empirical evidence from related research utilizing a HMD for CFF training that revealed an association between symptoms of simulator sickness and simulation exposure durations of at least 39 minutes (Champney, Lackey, Stanney, & Quinn, 2015). During the evaluation scenario of the second session, the SME interaction was restricted to that of the FDC role player, thus, the time in the system was determined by the participant's ability to complete the mission. The average time during the second session was 22 minutes, therefore, the trend toward reduced simulator sickness at the post measurement may be a consequence of the shorter exposure duration. Simulator sickness is a critical consideration for the design and development of emerging simulation technologies. Therefore, follow on experimentation should continue to track trends in simulator sickness in order to determine time in system thresholds.

CONCLUSION

This experiment represents an initial metrics and protocol assessment in a series of experiments planned to evaluate the training effectiveness of AV technologies applied to the CFF task domain. The foundational insight provided confirms the applicability and value of the selected assessment variables (i.e., performance metrics, self-efficacy, usability, immersion, etc.). The results also unfold empirical recommendations that may inform and improve the design of follow-on assessments. First, the sample population should be broadened to span the range of CFF expertise from novices to experts, which should assist in delineating the degree to which the AV training system, versus skill level, impacts the training experience and performance. Next, investigators should explore scalable performance measurement methods versus the binary standard, which will enable a more granular approach to performance diagnosis. Finally, comparisons of performance and subjective data between the CFFT-AV and other training systems will assist in determining how AV technologies rank in effectiveness and learner perceptions across the SBT paradigm. Overall, the collective lesson learned from these results and recommendations is that investigators must follow a dimensional approach to training effectiveness evaluation of SBT solutions for highly demanding tasks, such as CFF.

ACKNOWLEDGEMENTS

This research was sponsored by the U.S. Army Research Laboratory – Human Research Engineering Directorate Simulation and Training Technology Center (ARL HRED STTC), in collaboration with the Institute for Simulation and Training at the University of Central Florida. This work is supported in part by ARL HRED STTC contract W911QX-13-C-0052. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL HRED STTC or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

Champney, R., Lackey, S., Stanney, K., & Quinn, S. (2015). Augmented reality training of military tasks: Reactions from subject matter experts. *Proceedings from the 17th Annual International Conference on Human Computer Interaction* (HCII 2015).

Charlton, J. & Danforth, I. (2005). Distinguishing addiction and high engagement in the context of online game playing. *Computers in Human Behavior, 23*(3), 1531-1548.

Crovella, M.E. & Lipsky, L. (1997). Long-lasting transient conditions in simulations with heavy-loaded workloads. *Proceedings of the 1997 Winter Simulation Conference*.

Classen, S., Bewernitz, M., Shechtman, O. (2011). Driving simulator sickness: An evidence-based review of the literature. *The American Journal of Occupational Therapy, 65*(2), 179-188.

Fragomeni, G., Lackey, S.J., Champney, R., Salcedo, J.N., & Serge, S. (2015). Training effectiveness evaluation: Call for Fire Trainer – Augmented Virtuality (CFFT-AV). *Proceedings from the 17th Annual International Conference on Human Computer Interaction* (HCII 2015).

Gamito, P., Morais, D., Oliveira, J., Gamito, L., & Anastácio, M. (2006). Presence: Head mounted display vs. Translucid screen. *Annual Review of Cybertherapy and Telemedicine*, 4111-116.

Haque, S. & Srinivasan, S. (2006). A meta-analysis of the training effectiveness of virtual reality surgical simulators. *IEEE Transactions on Information Technology in Biomedicine, 10*(1), 51-58.

Hughes, C.E., & Stapleton, C.B. (2005). The shared imagination: Creative collaboration in Augmented Virtuality. Proceedings of the 11Th International Conference on Human-Computer Interaction.

Jackson, G.T., & McNamara, D.S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology, 105*(4), 1036-1049.

Jennett, C., Cox, A., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies, 66*(9), 641-661.

Kennedy, R., Lane, N., Berbaum, K., & Lilienthal, M. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203-220.

Kirkpatrick, D.L., & Kirkpatrick, J.D. (2006). *Evaluating Training Programs: The Four Levels* (3rd ed.). San Francisco, CA: Berrett-Koehler Publishers, Inc.

Lackey, S.L., Champney, R., Fragomeni, G., Salcedo, J.N., & Serge, S. (2015). A step toward data driven JFO human behavior models (Poster). *Proceedings of the 2015 Behavior Representation in Modeling and Simulation Conference (BRMIS 2015).* Washington DC.

Lackey, S.L., Salcedo, J.N., Matthews, G., & Maxwell, D. (2014). Virtual world room clearing: A study in training effectiveness. *Proceedings of the 2014 Interservice/Industry Training, Simulation, and Education Conference* (IITSEC 2014). Arlington, VA: NTSA.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems, 77*(12), 1321-1329.

Moss, J. D., & Muth, E. R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors, 53*(3), 308-319.

Oser, R.L., Gualtieri, J.W., Cannon-Bowers, J.A., & Salas, E. (1999). Training team problem solving skills: An event-based approach. *Computers in Human Behavior, 15*, 441-462.

Regenbrecht, H., Lum, T., Kohler, P., Ott, C., Wagner, M., Wilke, W., & Mueller, E. (2004). Using augmented virtuality for remote collaboration. *Presence: Teleoperators & Virtual Environments, 13*(3), 338-354.

Ryan, T. (2011, January-February). Keeping it Real: Don't Let Joint Fires Observer Skills Deteriorate. Retrieved from http://www.globalsecurity.org/military/library/report/call/call_11-35_ch15.htm.

Strensrud, B. S., Fragomeni, G., & Garrity, P. (2013). Autonomy requirements for virtual JFO training. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (I/ITSEC 2013). Arlington, VA: NTSA.

UNC Charlotte ROTC. (2010). Call for Fire: Key Points [PowerPoint Slides]. Retrieved from http://www.padocss.com/read~5454c29c335b7ba9628b456d.

U.S. Army. (1991). *FM 6-30: Tactics, Techniques, and Procedures for Observed Fire*. Washington, DC: Department of the Army.

U.S. Army PEO STRI (2012, November 26). Call for Fire Training (CFFT II Plus). Retrieved from http://www.peostri.army.mil/PRODUCTS/CFFT+/.

Van der Land, S., Schouten, A. P., Feldberg, F., Van den Hooff, B., & Huysman, M. (2013). Lost in space? Cognitive fit and cognitive load in 3D virtual environments. *Computers in Human Behavior, 29*(3), 1054-1064.

Wang, X., Day, E., Kowollik, V., Schuelke, M. J., & Hughes, M.G. (2013). Factors influencing knowledge and skill decay after training: A meta-analysis. In W. R. Arthur, E. Day, W. R. Bennett, A. M. Portrey (Eds.), *Individual and Team Skill Decay: The Science and Implications for Practice* (pp. 68-116). New York, NY, US: Routledge/Taylor & Francis Group.

Wray, R.E., Laird, J.E., Nuxoll, A., Stokes, D., & Kerfoot, A. (2005). Synthetic adversaries for urban combat training. *AI Magazine, 26*(3), 82-92.

Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). The factor structure of the presence questionnaire. *Presence, 14*(3), 298-312.

Zhang, P., Li, N., & Sun, H. (2006). Affective quality and cognitive absorption: Extending technology acceptance research. *Proceedings of the Hawaii International Conference on System Sciences*, pp. 207-217.