Large Scale Testing and Evaluation of Virtual Environments for Infantry Soldier Tasks Comparing Mental Effort for Live Versus Virtual Training Performance Assessments

Douglas Maxwell
Advanced Training Systems Division
U.S. Army Research Laboratory
Orlando, FL
douglas.maxwell3.civ@mail.mil

Jimmy Zheng
Institute for Simulation and Training
University of Central Florida
Orlando, FL
jimmyzheng@knights.ucf.edu

ABSTRACT

The United States Army has a significant investment dedicated to the use of virtual environments for infantry soldier skills training. Although there is a pervasive attitude within the acquisition community that a training system's graphics quality is the strongest indicator of training quality and utility. A literature review of performance assessment of infantry soldiers who have been trained using simulation-based training technology reveals very little data exists to quantify the return on this investment and validate this assumption. Further, the Government Accounting Office in 2013 and 2016 calls for better assessment of costs and training performance to properly assess the systems. Beginning in 2014, researchers at the U.S. Army Research Laboratory and the University of Central Florida designed and executed a large-scale study that collected performance data from over 50 squads of dismounted infantry soldiers. Multiple training treatments were provided to the squads, including a baseline treatment where they were provided only traditional live training, only virtual, and a treatment that combines live and virtual. This paper will discuss a comparison of the mental effort of soldiers trained with virtual means to soldiers trained with traditional live means. Further, the analysis will include the Situational Training Exercise (STX) performance assessments of the soldiers to attempt to determine if there are any correlations between the differences in the mental effort of the training treatments.

ABOUT THE AUTHORS

Dr. Douglas B. Maxwell is a Science and Technology Manager at the U.S. Army Research Laboratory-Human Research and Engineering Directorate (ARL-HRED). He is the director of the Military Open Simulator Enterprise Strategy (MOSES) project. Dr. Maxwell earned his Ph.D. in Modeling & Simulation.

Mr. Jimmy Zheng is currently a second-year graduate student at the University of Central Florida and is pursuing a PhD. in Industrial Organizational Psychology. His current research involves the use of technology for training, employee selection, and assessment. He is also involved in an Occupational Health Psychology lab that studies employee well-being and counterproductive work behavior.

Large Scale Testing and Evaluation of Virtual Environments for Infantry Soldier Tasks Comparing Mental Effort for Live Versus Virtual Training Performance Assessments

Douglas Maxwell
Advanced Training Systems Division
U.S. Army Research Laboratory
Orlando, FL
douglas.maxwell3.civ@mail.mil

Jimmy Zheng
Institute for Simulation and Training
University of Central Florida
Orlando, FL
jimmyzheng@knights.ucf.edu

INTRODUCTION

The current program of record for a game based virtual environment is the Virtual Battle Spaces 3 (VBS3) product. The VBS3 is based on a commercial gaming engine called ARMA3 developed by Bohemia Interactive. The VBS3 is widely used within the U.S. Army for various simulation-based training activities, including but not limited to infantry soldier skills training. The VBS3 is often criticized for being difficult to deploy and use, however, actual usage data is difficult to acquire. Due to licensing limitations, an additional game based virtual environment, MOSES (Military Open Simulator Enterprise Strategy) prototype.

The information contained in this paper is derived from field observations and data collection activities from March 2015 through December of 2016 at the 211th Florida Army National Guard Regional Training Institute's Basic Leader Course. The purpose of the data collection activities was to determine baseline training effectiveness comparison of simulation-based training methods versus physical walkthrough means. Soldier performance was assessed in situational training exercises in addition to the administration of questionnaires.

A research team composed of U.S. Army Research Laboratory (ARL) and University of Central Florida Institute for Simulation Technology (IST) personnel conducted a series of data collection activities by folding into the monthly Basic Leader Course training cycle conducted at the 211th RTI. Each month the team would travel to Camp Blanding, Florida and collect questionnaire responses from soldiers and performance evaluations from the course managers. By combining the team member's observations, the questionnaire data, and the performance evaluations, certain conclusions backed by statistically relevant data could be drawn.

This paper presents findings from a large-scale data collection activity. A background literature review revealed a paucity of data exists for infantry soldiers trained with game based virtual environments versus traditional live training means (Maxwell, 2015).

BACKGROUND

The United States Army has invested significant funding dedicated to the use of virtual environments (VEs) for training infantry soldier skills. There is a pervasive attitude in the acquisition community that a simulation-based training (SBT) system's graphics quality are the strongest indicators of utility and training quality (Stevens, 2014). Very little data exists to quantify the return on investment (ROI) provided by these training systems (Bell, Kanar, & Kozlowski, 2008). There is also a lack of formal methodologies for the identification of where in the training cycle these technologies belong as well as which training tasks they should be applied (Bowers et al., 2013). The United States Government Accountability Office issued a report in August of 2013 which calls for better assessment of performance and accounting of costs to accurately assess SBT systems throughout the United States Army and Marine Corps (Pickup, 2013).

To further complicate matters, the lack of formal requirements and performance measurement methodologies has led to a fracturing of the training space within the United States military that utilizes game-based virtual environments (GBVE). Although there is a GBVE training system listed as the program of record called Virtual Battlespace, it is limited for specialized training needs of some organizations. Pockets of innovation and product development in recent years have resulted in numerous training systems specializing in different utilization such as education and military applications (Buede, Deblois, Maxwell, & Mccarter, 2013).

The United States Army requires a mechanism for properly assessing the performance of infantry soldiers who have been trained using virtual simulations to establish statistically significant differences (if any) for comparisons to traditional training methods. This research initiative was conducted through Cooperative Agreements (CA) #W911NF-14-0012 and #W911NF-15-0004 between the United States Army Research Laboratory and the University of Central Florida. These CAs were created to facilitate the investigation of training effectiveness of operationally relevant tasks in a VE as compared to traditional classroom and live training. The desired outcome of this work is to establish a methodology for quantitatively defining the training effectiveness differences between traditional and virtual methods, and acquiring data through field experimentation to apply the methodology.

A literature review has revealed a lack of knowledge surrounding the efficacy of the practical application of virtual world technology for infantry soldier training, specifically ground combat skills training such as room clearing and reaction to contact (Lackey, Salcedo, Matthews, & Maxwell, 2014). Due to the current subjective nature of gauging training effectiveness of VEs, it is difficult to calculate a ROI (Maxwell, 2016). Lastly, it is difficult to determine comparisons of knowledge transfer between traditional and virtual training activities for ground combat skills.

This research effort represents years of organization, data collection, and analysis efforts. The research was possible only through the cooperation of the 211th RTI command staff and course managers, the University of Central Florida, and the U.S. Army Research Laboratory. As of this writing, data collected from 58 squads of soldiers have been examined and evaluated and analyzed.

INFANTRY SOLDIER SKILLS ASSESSMENT

Observations in this paper derived from data collection events performed in early 2015 through the end of 2016 at the Camp Blanding Regional Training Institute at the 211th Florida Army National Guard. The Basic Leadership Course (BLC) managers worked closely with the ARL/UCF research team to determine how best to use it for a comparison study. This study was extremely fortunate to find an accommodating unit to allow for repeated observation and even adopt suggested adjustments to the course.

A between-treatments experiment was designed to make several comparisons of the training effectiveness of soldiers provided a virtual training method versus traditional training means, virtual training method versus a blended virtual and traditional training method, and lastly, traditional training compared to a blended virtual and traditional training method. The period of instruction for this BLC is 20 days, with days one through 17 consisting of all soldiers provided the same classroom training with slides and instruction subject matter experts. Normally, day 18 is reserved for practical exercises and four hours is provided for simulator time. Traditionally, practical exercises are provided in the form of instructor led “walkthroughs” performed at a wood line or other designated area serving as a live proxy for the operational environments discussed in class. Virtual “walkthroughs” are provided at simulation centers with laptops where the four major tasks are posed to the soldiers by instructors and they practice reacting to the situational stimulus. The simulation can pose situations such as indirect fire, improvised explosive device detection, land navigation to rally points, and more. The 211th RTI uses the U.S. Army’s VBS3 product for this simulation and an additional classroom has been prepared with access to the MOSES prototype.

The formal assessment of the squad’s performance is done on days 19 and 20 using on-site situational training exercise (STX) lanes. For this experiment, and adjustment was made to the period of instruction (POI) such that the class was separated into two groups and provided with different walkthrough training treatments. The control group was sent to a wooded area and an instructor provided practical exercises with guided instruction (Fig. 1.). This control group represents a traditional method of providing the practical instruction.

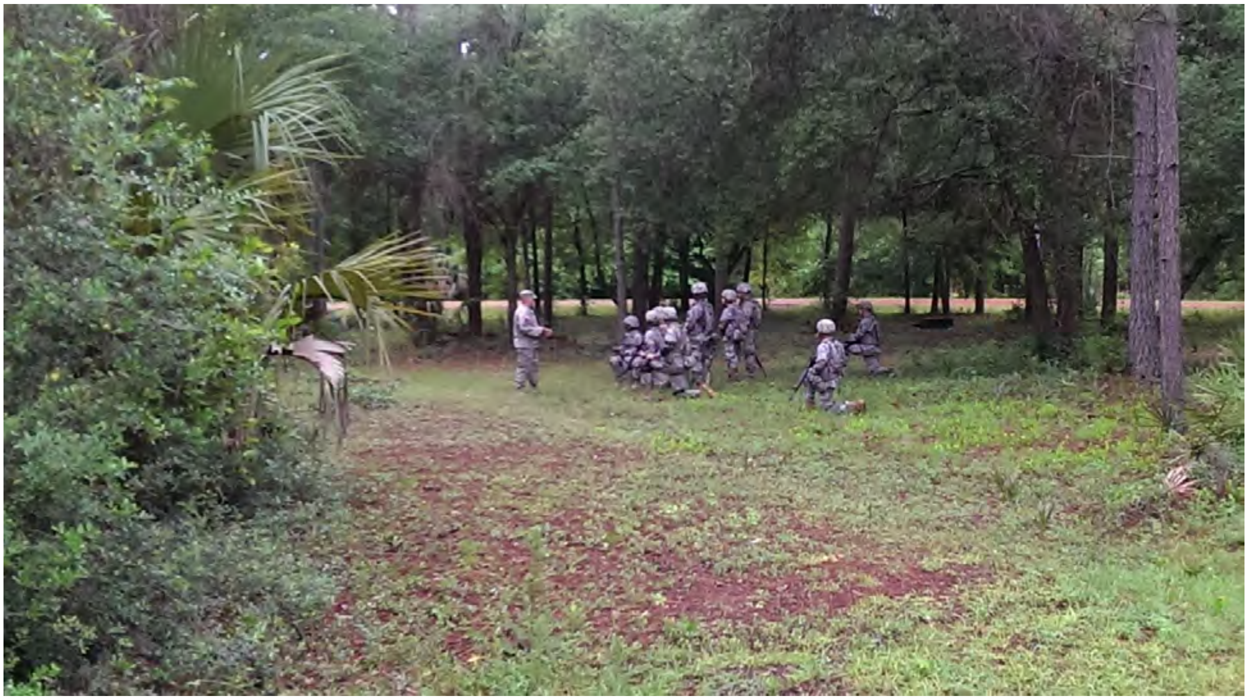

Fig. 1. Small Group Leader Providing Guided Instruction

The experimental groups were provided a training treatment that included simulation-based training in a classroom setting and used game-based virtual environments such as the VBS3 suite (Fig. 2.) and the MOSES prototype (Fig. 3.). This paper treats the virtual treatment as a generic technology. As more data is collected, we will begin to have enough statistical significance to be able to make simulator technology comparisons of different products and technologies effect on soldier performance. The scenario provided within the VBS3 was created by onsite contractors and approximated the physical STX lanes the soldiers would encounter the following day during their performance evaluations. The scenario provided within the MOSES was created by UCF/IST graphic artists and approximated the physical STX lanes.

The ARL/UCF research team was allowed unrestricted access to the course managers, classrooms, battle training center, practices areas, and STX lanes. The accommodations made by the 211th allowed for a deep examination of the course and the affect a simulation-based training system had on the performance of infantry soldiers. The course managers were provided with periodic updates and briefed regularly on the progress of the data collection. The course managers also solicited advice from the research team on how to improve the use of the VBS3 system and played an integral role in the experimental design of the follow-on study scheduled to start in the Spring of 2017.

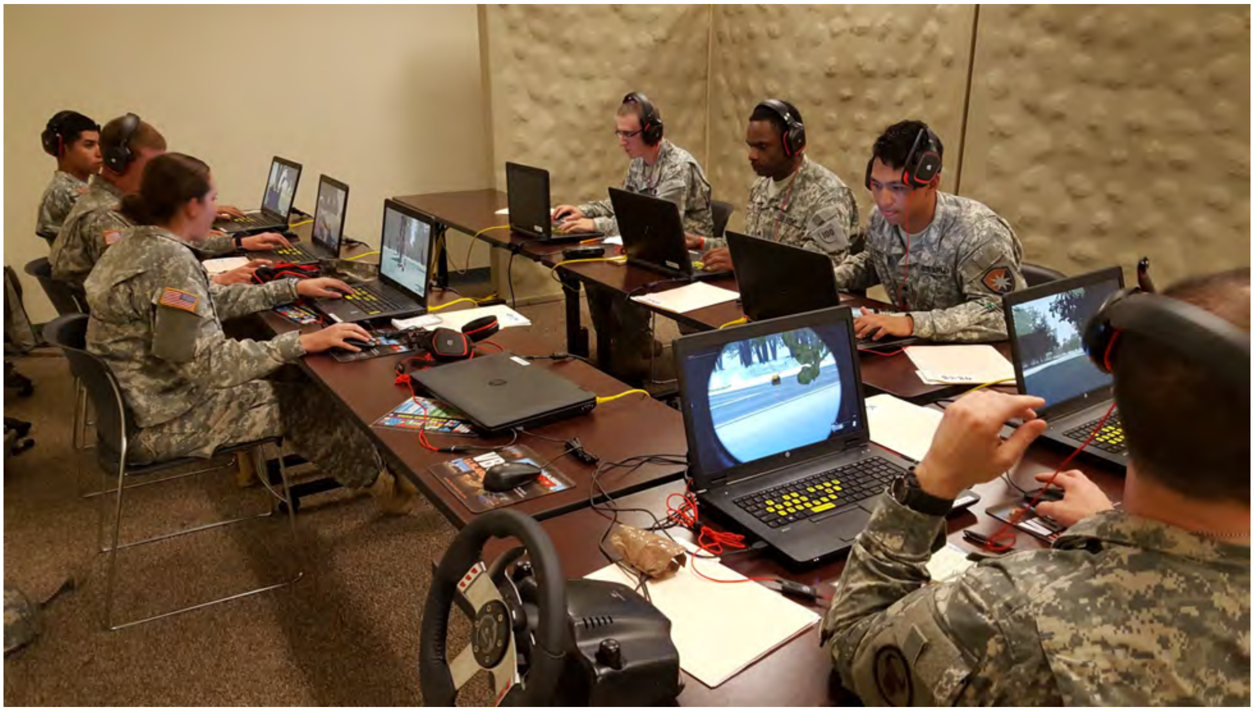

Fig. 2. Virtual Battle Spaces 3 Application in Use

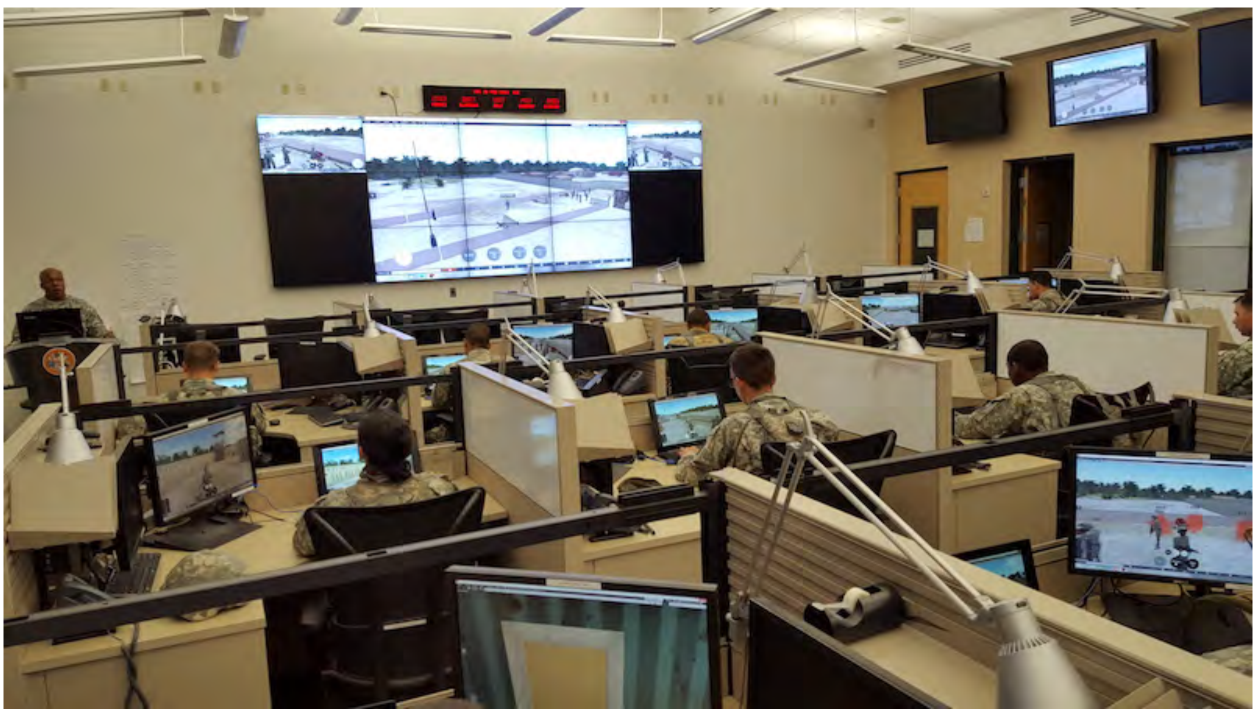

Figure 3. MOSES Application in Use

METHOD

The participants for this experiment were comprised of 573 Florida Army National Guard soldiers (59 squads). There were 451 males and 122 females. The age of participants ranged from 19 to 48 years (M = 27.6, SD = 5.3). The participants were recruited through the Basic Leadership Course coordinated with the Army post training personnel and each squad was randomly assigned to a condition upon arrival. Due to the variability of the class sizes and available course managers, the composition of the squads was not always uniform. Typically, a squad is composed of 9 people; one squad leader, two fire team leaders, two SAW gunners, two grenadiers, and two riflemen. Sometimes extra riflemen were added to each squad if the class size had an unusual number of soldiers in it.

The experiment utilized a between subject design that included one independent variable, the training treatment. Dependent variables included performance and survey responses.

The period of instruction (POI) for this course is 20 days. For days 1-17, the soldiers receive the same classroom-based training including PowerPoint slides and SME instruction. Typically, day 18 is reserved for practical exercises with a four-hour block allocated for scenario-based training. The practical exercises consist of "walkthroughs," where the four major tasks are posed to the soldiers and they are given the opportunity to practice responding to the task. The 211th RTI uses the United States Army's Virtual Battlespace 3 (VBS3) simulation platform for training. On days 19 and 20, a formal assessment of the squad's performance is evaluated during an on-site situational training exercise (STX).

For this experiment, an adjustment was made to the POI such that each month, the class was separated into multiple groups and provided with different walkthrough training treatments. This control group represented a traditional method of providing practical instruction, the live walkthroughs. The practical instruction comprised of sending the control group into the near-by wooded areas where an instructor provides guided instruction during the practical exercises for four hours. The baseline training treatment had 169 soldiers (16 squads). During the four hours not in the classroom, the soldiers were given non-training activities such as logistical preparation for the following day's activities.

The "virtual only" experimental group received the training treatment in a computer lab using the VBS3 suite. The VBS3 scenarios were developed by onsite contractors who replicated the STX lanes the soldiers would encounter the following day during their performance evaluations. The soldiers who received only practical exercises in simulation form spent 4 hours in a classroom with the VBS3 suite. Space is limited in the classroom, so some of the soldiers were given a morning shift and some were given an afternoon shift. As with the baseline treatment, the four hours not in the classroom the soldiers were given non-training activities such as logistical preparation for the following day's activities. The virtual only treatment had 174 soldiers (15 squads).

For training treatment three, soldiers provided both live walkthroughs and exposure to virtual training environments. The period of instruction for each exposure was four hours and the soldiers were provided with guided instruction from the course managers.

Table 1. Training Treatments, Number of Soldiers and Squads

Training Treatment	n Soldiers	n Squads
1 – Baseline (Live Walkthroughs)	169	16
2 – Virtual Only	174	15
3 – Live and Virtual	230	28
Total	573	59

On day 18 of the POI, all soldiers enrolled in the BLC were assembled and provided a brief describing the experiment. The soldiers were given the opportunity to ask questions pertaining to the experiment before signing consent forms. The UCF/ARL IRB reviewed the consent form to ensure the study had minimal risk to the soldiers. After the soldiers signed the consent forms, they completed a series of pre-experimental questionnaires.

Following the walkthroughs on day 18, the two treatment groups assembled at the STX lanes (on day 19) to participate in the live simulation that evaluated the performance of the training tasks.

Table 2 shows the original evaluation task list used in the course for team performance evaluations. To gather the data from the BLC, it was necessary to adjust the original course rubric so that a more meaningful comparison could be made between the control and virtual training treatment. As with most Army training norms, the BLC training also relied on a "Go/No-Go" performance evaluation metric.

This research focus seeks to determine the applicability of specific infantry soldier skills against different training treatments. The current rubric indicates differences between the training treatments, but a new rubric is required to provide a comparison of how much one treatment differed from another. Each major training task was divided into subtasks and the assessment performed utilized a four point Likert scale (Coltekin, Heil, Garlandini, & Fabrikant, 2009). A four point Likert scale provides an opportunity to make a choice by eliminating the midpoint responses. This research expands the rating categories from two (i.e., GO/NO-GO) to four (i.e., needs improvement, adequate, successful, and excels) which enables the research team, as well as the course cadre, to gain greater insight into whether the preceding training condition influenced trainee performance. Coded categorical data can be treated as numerical and lends itself to deeper analysis if the optimal number of categories are employed (Maxwell, Stevens, & Maraj, 2016; Maxwell, 2015). For this study, four categories of rated performance were created through a questionnaire and used to not overload cadre rating requirements (i.e. performing the actual evaluations); while providing the research team with quantifiable data for analysis. Further, subjectivity in the evaluation was reduced by decomposing the training tasks to the subtask level, thereby allowing the cadre to increase their objectivity of ratings in the performance evaluation at each atomic step. Table 3 shows the adjusted rubric the ARL/UCF team provided to the SGLs for use in their final squad performance evaluations.

Table 2. Task List for Collective Performance Assessment for 211th FLANG RTI BLC

Task
1. React to Indirect Fire while Dismounted
2. React to Improvised Explosive Device
3. React to Near Ambush
4. React to Far Ambush

The SGLs assessed the squads (or groups) per a new rubric in real-time. A total of 59 squads were evaluated and over 570 soldiers provided questionnaire input. The squads followed the experimental protocols explicitly, yielding data for 16 squads exposed to the control treatment and 15 squads with the virtual treatment and 28 squads exposed to the blended live and virtual training treatment.

By applying the new rubric, many data points were generated allowing the team to apply parametric statistics to the survey data and calculate analysis of variables on the squad performance data. Implementation of the new rubric was instrumental for gathering expanded data points to calculate comparisons between the two treatment groups.

Table 3. Indirect Fire Task with Subtasks and Revised Rubric

Task	**Assessment**			
	Needs Improvement	Adequate	Successful	Excels
1. React to Indirect Fire while Dismounted				
1.1 Shout "Incoming" in a loud, recognizable voice.				
1.2 React to the Instruction of the leader and look for guidance.				
1.3 Seek nearest cover.				
1.4 Assess Situation				
1.5 Report situation to leader				
1.6 Continue Mission				

RESULTS

Data was obtained from the NASA-TLX mental effort scores and course manager's performance ratings of the soldiers in the STX lanes.

Mental Effort

A one-way between-groups ANOVA was conducted to compare the effect of training condition on mental demand experienced by soldiers in the virtual, live, and mixed conditions. We found a significant effect of training condition on mental demand for the three training conditions $F(2,396) = 4.48$, $p = .012$. A post hoc Tukey test was conducted to compare training conditions (virtual vs. live, virtual vs. mixed, and live vs. virtual). We found the mental demand score for the live condition ($M = 50.80$, $SD = 23.04$) did not differ significantly from the mental demand score for the virtual condition ($M = 53.99$, $SD = 25.49$, $p = .54$) or mixed condition ($M = 45.89$, $SD = 21.48$, $p = .24$). However, there was a significant difference in mental demand scores between the mixed condition and the virtual condition ($p = .009$).

They key learning objectives for this training were for soldiers acting in a squad leadership role to be able to recognize the type of contact they encountered and to respond per the doctrine outlined in the classroom. While this has traditionally been treated as a kinetic task, an examination of the steps required to successfully complete the task reveal the requirement to understand environmental cues. The faithful replication of the environmental cues in the simulated training areas, whether virtual or live was critical to the success of the training. For example, a factor such as the sound of the fire to indicate the type of weaponry used in an ambush indicates small arms or machine gun fire. In the live walkthroughs, proxy sounds were used such as clapping two pieces of wood together by the course manager. In the virtual walkthroughs, replay of recordings of proper sounds were used. It would seem to give the virtual treatment an advantage.

However, the direction of fire is also a critical component to the decision-making process. The course managers in the live walkthroughs easily replicate this by generating proxy gunfire sounds from the wood line. This is problematic in the virtual walkthroughs. While positional audio is possible inside the virtual simulators, the soldiers are required to wear headphones to take advantage of the system. Further, the pace of the scenarios often seemed to inhibit the use of the headphones and the soldiers were often observed taking them off so they can hear each other's commands. The use of radio functions within the simulator were usually ignored as they simply shouted orders at each other within the classroom.

These and many more tradeoffs were recognized and were the reason for the mixed virtual and live training treatments. By providing a mix of both treatments, it was hypothesized the strengths of exposure to both live and virtual training would exercise both kinetic and cognitive skills required to successfully complete the task. The live condition represents the baseline mental effort for the training. The comparison of mental effort between live and virtual condition did not yield significant differences and the mixed condition yielded the counterintuitive significant result of lowest overall reported mental effort.

Performance Comparisons

A Mann-Whitney U test was used to compared squad STX collective task performance (IDF task, IED task, React Near Ambush task, React Far Ambush task) between multiple training conditions (virtual vs. live, virtual vs. mixed, live vs. mixed).

A comparison test was performed between the collective task performance for the virtual and live training conditions. The IDF task performance comparison was significant ($U = 43.5$, $p = .012$), squads trained in the live condition did better ($Mdn = 2.83$) than squads trained in the virtual condition ($Mdn = 2.50$). We also found IED task performance was higher for squads trained in the live condition ($Mdn = 2.75$) than squads trained in the virtual condition ($Mdn = 2.33$) but was nonsignificant ($U = 73.5$, $p = .161$). React Near Ambush task performance was higher for squads trained in the live condition ($Mdn = 2.70$) than squads trained in the virtual condition ($Mdn = 2.40$) but was nonsignificant ($U = 86.5$, $p = .416$). React Far Ambush task performance was higher for squads trained in the live condition ($Mdn = 2.33$) than squads trained in the virtual condition ($Mdn = 2.25$) but was nonsignificant ($U = 72.5$, $p = .244$).

Although only one of the tasks resulted in significant differences between the live and virtual treatment, the descriptive statistics show that the scores were lower across the board for soldiers trained using only the virtual walkthroughs. This indicates the virtual training is not properly replicating the operational environment, not properly replicating the training tasks, not properly replicating the interpersonal communication mechanisms needed to operate as a collective, or some combination of all.

The next test compared task performance between the virtual and mixed training conditions. The IDF task performance comparison was significant ($U = 44.0$, $p = .000$), squads trained in the mixed condition did better (*Mdn* = 3.09) than squads trained in the virtual condition (*Mdn* = 2.50). Additionally, the IED task performance comparison was significant ($U = 81.5$, $p = .022$), squads trained in the mixed condition did better (*Mdn* = 2.90) than squads trained in the virtual condition (*Mdn* = 2.33). The React Near Ambush task performance was higher for squads trained in the mixed condition (*Mdn* = 3.00) than squads trained in the virtual condition (*Mdn* = 2.40) but was nonsignificant ($U = 116.0$, $p = .251$). React Far Ambush task performance was higher for squads trained in the mixed condition (*Mdn* = 2.50) than squads trained in the virtual condition (*Mdn* = 2.25) but was nonsignificant ($U = 91.0$, $p = .072$).

The virtual versus mixed comparison shows that the introduction of live exposure to the soldiers provided improvement over virtual only treatments. Specifically, the identification and reaction to indirect fire and improvised explosive devices. It is hypothesized the field of view and resolution of the simulator (in this case a 15" laptop screen) was not sufficient to replicate the actual environment to provide an adequate replication of the real IED identification task. With respect to the indirect fire, the direction of fire was not easily identified and it was consistently observed by the research team that the students were unable to properly issue orders to their squad for which direction to go. Similar observations were made for the near and far ambushes, however, the simulator could provide more visual cues and therefore the students could more easily react appropriately.

Lastly, we compared task performance between the live and mixed training condition. IDF task performance for squads trained in the mixed condition was higher (*Mdn* = 3.09) than squads trained in the live condition (*Mdn* = 2.83) but not significant ($U = 85.0$, $p = .096$). IED task performance was higher for squads trained in the mixed condition (*Mdn* = 2.90) than squads trained in the live condition (*Mdn* = 2.75) but not significant ($U = 115.0$, $p = .378$) React Near Ambush task performance was higher for squads trained in the mixed condition (*Mdn* = 3.00) than squads trained in the live condition (*Mdn* = 2.70) but not significant ($U = 118.0$, $p = .437$). React Far Ambush task performance was higher for squads trained in the mixed condition (*Mdn* = 2.50) than squads trained in the live condition (*Mdn* = 2.33) but not significant ($U = 101.5$, $p = .396$).

The descriptive statistics for the performance scores for soldiers with mixed treatment were higher for the soldiers with live only exposure. However, the lack of significance in the performance scores indicates the virtual training at little effect on the overall outcome.

FUTURE WORK

A baseline of indicators for training effectiveness of traditional and virtual simulators for infantry soldier skills training. This paper discusses the virtual treatment in generic terms, however, two different simulators were used. As more data is collected, enough statistical power can be established to look at differences in virtual trainers. The scenarios and actions within the scenarios were replicated and deemed equivalent by the course managers.

More factors need to be investigated that may influence overall performance. In the future, modifications to the user interfaces and scenario composition will be added to the training treatments and tested for effect on performance. Currently, the virtual systems are cumbersome and soldiers have been observed becoming frustrated with the operation of the simulators. Additional features need to be either added or augmented in the simulators such as better directional audio.

Augmented and virtual reality hardware is becoming more and more cost-effective. By providing a wider field of view, the operational environment of the infantry soldier may soon be able to be more accurately presented. Testing and evaluation of these systems needs to be performed before acquisition and deployment.

ACKNOWLEDGEMENTS

The U. S. Army Research Laboratory and the University of Central Florida research team would like to express our profound gratitude to the dedicated men and women of the 2/124th and 211th without whom this work could not have been performed. We thank you for your service to our country.

LIST OF ACRONYMS

BLC	Basic Leadership Course
IDF	Indirect Fire
IED	Improvised Explosive Device
MOSES	Military Open Simulator Enterprise Strategy
NASA-TLX	National Aeronautics and Space Administration - Task Load Index
POI	Period of Instruction
RTI	Regional Training Institute
STX	Situational Training eXercise
VBS3	Virtual Battle Spaces ver. 3

REFERENCES

Bell, B. S., Kanar, A. M., & Kozlowski, S. W. J. (2008). Current issues and future directions in simulation-based training in North America. *The International Journal of Human Resource Management*, *19*(January 2015), 1416–1434. http://doi.org/10.1080/09585190802200173

Bowers, C. a., Serge, S., Blair, L., Cannon-Bowers, J., Joyce, R., & Boshnack, J. (2013). The Effectiveness of Narrative Pre-Experiences for Creating Context in Military Training. *Simulation & Gaming*, *44*(4), 514–522. http://doi.org/10.1177/1046878113475341

Buede, D., Deblois, B., Maxwell, D., & Mccarter, B. (2013). Filling the Need for Intelligent, Adaptive Non-Player Characters Filling the Need for Intelligent, Adaptive Non-Player Characters. In *Interservice/Industry Training Simulation and Education Conference (I/ITSEC)* (pp. 1–10).

Coltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-movement Analysis. *Cartography and Geographic Information Science*, *36*(1), 5–17.

Lackey, S. J., Salcedo, J. N., Matthews, G., & Maxwell, D. B. (2014). Virtual World Room Clearing : A Study in Training Effectiveness. *Interservice/Industry Training, Simulation, and Education Conference*, (14045), 1–11. Retrieved from www.iitsec.org

Maxwell, D. B. (2015). *Gauging Training Effectiveness of Virtual Environment Simulation Based Applications for An Infantry Solider Training Task*. University of Central Florida. Retrieved from http://etd.fcla.edu/CF/CFE0005840/MAXWELL_Dissertation_10_JUN_2015.pdf

Maxwell, D. B. (2016). Application of virtual environments for infantry soldier skills training: We are doing it wrong. *Human Computer Interaction International*, *9740*, 424–432. http://doi.org/10.1007/978-3-319-39907-2_41

Maxwell, D., Stevens, J., & Maraj, C. (2016). Alternate rubric for performance assessment of infantry soldier skills training. *Human Computer Interaction International*, *9744*, 34–45. http://doi.org/10.1007/978-3-319-39952-2_4

Pickup, S. (2013). *ARMY AND MARINE Better Performance and Cost Data Needed to More Fully Assess Efforts*. Retrieved from http://www.gao.gov/assets/660/657115.pdf

Stevens, J. A. (2014). *a Program Manager'S Dilemma: Measuring the Effect on Performance of Different Visual Modalities in Mixed Reality Aerial Door Gunnery*. University of Central Florida. Retrieved from http://stars.library.ucf.edu/etd/4557/