Measuring Training Effectiveness of Lightweight Game-based Constructive Simulation

Dr. Jonathan Stevens
Dr. Stephen R. Serge
Mr. Dean Reed
University of Central Florida (UCF)
Orlando, FL
jonathan.stevens@knights.ucf.edu
sserge@ist.ucf.edu
dreed@ist.ucf.edu

Ms. Latika Eifert
Army Research Laboratory (ARL)
Orlando, FL
latika.eifert@us.army.mil

Dr. Boris Stilman
University of Colorado Denver
STILMAN Advanced Strategies
Denver, CO
boris@stilman-strategies.com

Dr. Oleg Umanskiy
STILMAN Advanced Strategies
Denver, CO
oleg@stilman-strategies.com

ABSTRACT

The U.S. Army continues to employ constructive and game-based simulation for training. While both classes of simulation have been found to lower the cost of training, it is still unknown whether or not these classes are actually effective training mechanisms. The Linguistic Geometry Real-time Adversarial Intelligence and Decision-making (LG-RAID) simulation is a lightweight, game-based, constructive simulation that exploits novel game theory to create intelligent, predictive and tactically-correct Courses of Action (COAs) for exercise participants at the company echelon and below. The primary goal of this study was to examine the training effectiveness and usability of the U.S. Army's LG-RAID simulation in an operationally relevant environment. A secondary objective of this study was to assess both the usability and functionality of the simulation in order to improve the technology through future design recommendations. Qualified Soldiers were randomly assigned to one of two training treatments (LG-RAID or a traditional planning method) and tasked to develop, plan and brief a tactically sound operational mission in order to empirically assess the training effectiveness of LG-RAID. The independent variable was training treatment. Dependent variables included performance and individual survey responses. Experimentation was conducted at Fort Benning, GA and performance was evaluated by accredited Army instructors. Results of this study indicate that LG-RAID shows promise as an effective training simulation tool when compared to the baseline condition.

ABOUT THE AUTHORS

Dr. Jonathan Stevens is a Research Scientist with the Institute for Simulation and Training (IST) at the University of Central Florida (UCF), specializing in training simulation research. Dr. Stevens is a retired Lieutenant Colonel of the United States Army with over 22 years of military experience as both an Infantry and Acquisition Corps officer. He received his Ph.D. in Modeling and Simulation from UCF.

Latika (Bonnie) Eifert is a Science and Technology Manager at the U.S. Army Research Laboratory-Human Research and Engineering Directorate, Simulation and Training Technology Center (ARL-HRED STTC) located in Orlando, Florida. Ms. Eifert manages several projects associated with simulation and training and also supports the Defense Advanced Research Project Agency (DARPA) by managing various research program efforts. She received her M.S. in Computer Engineering from University of Central Florida in 2003.

Dr. Stephen R. Serge is a Research Associate with the Institute for Simulation and Training (IST) at the University of Central Florida (UCF), specializing in user research, usability, and training design and effectiveness. Dr. Serge has been directly involved with game and simulation-based training research and design for over five 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.

Dean Reed is a Senior Associate for Simulation with the Institute for Simulation and Training (IST) of the University of Central Florida (UCF). Mr. Reed is a veteran of the U.S. Army, leads a team of developers at IST and has been using variations of Semi-Automated Forces (SAF) software throughout his 15-year career. He received his B.S. in Computer Science from UCF in 2000.

Dr. Boris Stilman is the Chairman and CEO at STILMAN Advanced Strategies (STILMAN) and Professor of Computer Science at the University of Colorado Denver. Dr. Stilman leads STILMAN since its foundation in 1999. All the software applications including LG-RAID developed at STILMAN are based on Linguistic Geometry, a new type of game theory in Artificial Intelligence, which has been originated and further developed by Dr. Stilman since 1972. He has published over 200 research papers.

Dr. Oleg Umanskiy is the Chief Software Architect at STILMAN Advanced Strategies (STILMAN) and leads the software development effort of the LG-RAID simulation. Oleg has over 18 years of software design, development and research experience, and has been the architect and the leading developer of the LG-RAID simulation for STILMAN. He received special recognition for his work from the Defense Advanced Research Projects Agency (DARPA). He received his Ph.D. in Computer Science and Information Systems from University of Colorado Denver in 2015.

.

Measuring Training Effectiveness of Lightweight Game-based Constructive Simulation

Dr. Jonathan Stevens
Dr. Stephen R. Serge
Mr. Dean Reed
University of Central Florida (UCF)
Orlando, FL
jonathan.stevens@knights.ucf.edu
sserge@ist.ucf.edu
dreed@ist.ucf.edu

Ms. Latika Eifert
Army Research Laboratory (ARL)
Orlando, FL
latika.eifert@us.army.mil

Dr. Boris Stilman
University of Colorado Denver
STILMAN Advanced Strategies
Denver, CO
boris@stilman-strategies.com

Dr. Oleg Umanskiy
STILMAN Advanced Strategies
Denver, CO
oleg@stilman-strategies.com

INTRODUCTION

The purpose of this research was to examine the training effectiveness of the Linguistic Geometry Real-time Adversarial Intelligence and Decision-making (LG-RAID) simulation tool in an operationally relevant environment. LG-RAID is a light-weight software tool that generates intelligent, predictive and tactically-correct Courses of Action (COAs) for constructive exercise participants at the company echelon and below. LG-RAID employs Linguistic Geometry (LG), a theory derived from both Artificial Intelligence (AI) and Game Theory.

The primary objective of the experiment described in this paper was to empirically assess the training effectiveness of LG-RAID using a qualified population of Soldiers, with a secondary objective being the assessment of the usability and functionality of the simulation in order to improve the technology.

There are three classes of simulation used for training: live, virtual and constructive (Hodson & Hill, 2013). In the past decade, a fourth class has been added: gaming (Roman & Brown, 2008). Live simulations comprise real people operating real systems while virtual simulations encompass real people operating simulated systems in simulated environments. Constructive simulations involve simulated people operating simulated systems. The term gaming refers to the employment of interactive, computer-based applications used for training purposes (Bergeron, 2006), generally characterized by their low overhead and cost. When the four classes overlap each other in training (i.e. employment of more than one class in a training exercise), this is referred to as "blended training".

One of the major advantages of employing virtual and/or constructive simulation in training is its cost advantages (Orlansky, et al., 1994), especially when compared to live training. Constructive simulation, in particular, has been found to be a significant cost-saving option (Riecken, et al., 2013), particularly considering that today's constructive simulations can employ thousands of entities simultaneously in the simulation.

While LG-RAID has been proven to be an effective simulation driving other constructive simulations (Stevens, Eifert, Reed, Diaz, & Umanskiy, 2014), this technology can also be utilized as a game-based trainer. However, no empirical assessment has been conducted on its training effectiveness and usability while employed in a stand-alone manner, similar to a game. In fact, minimal empirical evidence exists regarding the effectiveness of game-based training (Whitney, Tempby, & Stephens, 2014) (Sotomayor & Proctor, 2009).

LG-RAID remains an Army science and technology effort even after several years of capability development. Recent Army Research Laboratory (ARL) development effort has focused on improving the LG-RAID capability for the end-user by adding critical planning capabilities to a user-friendly and intuitive interface. Additional development has focused on migrating this simulation capability to a cloud-based service. As LG-RAID's capability has matured, it is now imperative that this critical technology be transitioned to an Army Program of Record (PoR) in order for Soldiers to leverage this capability. In order to facilitate that transition, an empirical assessment of the technology's training effectiveness, as a game-based trainer, was conducted. This study also examined the functionality and usability of the simulation, the results of which will be published in the future.

This collaborative investigation builds upon previous simulation-based training analysis between ARL and the University of Central Florida (UCF) Institute for Simulation and Training (IST). This study represents the first phase in a multi-phase series of experiments designed to investigate the training effectiveness and usability of LG-RAID for Soldier training.

BACKGROUND

LG RAID Background

LG-RAID is based on LG, a theory of Abstract Board Games in AI developed since 1972 (Stilman, 2000) (Stilman, Yakhnis, & Umanskiy, 2010). This approach is scalable and allows for the generation of winning strategies for real world systems in real time or near real time. This scalability is based on the avoidance of tree-based searches typical to other gaming systems, leading to the avoidance of combinatorial explosion that prevents scalability of those systems. In LG, the analysis (a tree-based search within the game state space) is replaced by the synthesis (construction of strategies within the Abstract Board).

LG-RAID development started as part of the Defense Advanced Research Projects Agency (DARPA) RAID program in 2004 with a goal of designing a decision support aid for lower tactical echelons with little computational support (Stilman, Yakhnis, & Umanskiy, 2007). Currently the LG-RAID simulation is capable of providing a low-overhead COA generation and wargaming capability that can be utilized for mission planning, rehearsal, and execution. Within the scope of the current development, these same capabilities are being leveraged to provide a light weight capability for simulation-based training.

Within a military classroom environment, the student can employ LG-RAID to evaluate and visualize alternative COAs for the military missions presented by the instructor. Using the web-based User Interface, the user can input mission data such as task-organized friendly forces (e.g., units, equipment, ammo loads, and positions), scheme of maneuver as an execution matrix, command and control (C2), mobility and fire control graphics, and anticipated enemy forces or plans. Based on such student COAs, LG-RAID automatically executes the scenario as a faster than real-time simulation, generating tactically valid actions for both friendly and enemy forces for the entire mission duration, without additional input from human operators or detailed scripting. The results of the simulation run are presented to the student as an animated movie showing likely movements and actions for all entities over the desired time horizon, as well as statistical output of expected casualties and ammunition expenditure. The user can then execute alternative COAs for the same mission to qualitatively and quantitatively compare the likely outcomes of such different strategies. The instructor can further challenge the student by executing the student’s BLUFOR COA against an unexpected and challenging OPFOR COA. Such an interactive and visual approach is intended to stimulate the reasoning about the tactical situation and understanding of the battlefield dynamics, leading the student to develop a more comprehensive and judicious COA for the training missions and enhance his/her mission planning proficiency.

Game-Based Training (GBT)

It has become increasingly common to observe more game-like qualities (i.e., scores and points, earning trophies/rewards, avatar personalization, etc.) incorporated into computer-based simulation training systems (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). The smaller-scale hardware and development requirements, mobile and PC-based platforms, and motivational or rewarding characteristics typically associated with games has led to an increased interest in utilizing GBT (Belanich, Orvis, & Sibley, 2013; Proske, Roscoe, & McNamara, 2014). While research has shown GBT as a viable option within the larger scope of training and learning, the effectiveness of these systems over more traditional methods of training has been inconsistent.

Findings from current GBT effectiveness research have found GBT outcomes similar to more traditional styles of training, such as classroom-based lectures and virtual simulation-based training (Proske, Roscoe, & McNamara, 2014; Gega, Norman, & Marks, 2007; Jackson & McNamara, 2013) This may be due to eclectic use of different types of games and methods of instructional development, leading to different results in the available research studies (Serge, 2014). In most instances, it appears that games for training are most beneficial when used in combination with other traditional forms of instruction (Sitzmann, 2011). This means that trainees receive some type of review or instructional session prior to interacting with the game, which serves as a supplemental resource for training. Researchers reporting on those approaches have found that reinforcement of training material and the ability to practice in the game has been beneficial when supplemental guidance is provided in the gaming environment (Serge, Priest, Durlach, & Johnson, 2013; Weiner, et al., 2011). Furthermore, other researchers have suggested that individuals with higher gaming proficiency tend to perform better on tasks executed within the gaming environment as a result of training (Keebler, Jentsch, & Schuster, 2014; Richardson, Powers, & Bousquet, 2011). The existence of a number of studies that find comparative results between traditional methods of training and GBT (Proske, Roscoe, & McNamara, 2014; Sitzmann, 2011) reveals that game-specific features are capable of promoting high levels of effective learning in simulated training systems.

At first glance, it not may seem beneficial to include game-like attributes to simulation-based trainers, so long as results are comparable to traditional methods of training. However, the game-like features of GBT also serve to increase user engagement and motivation for learning and interacting with the system versus their non-game counterparts (Garris, Ahlers, & Driskell, 2002; Girard, Ecalle, & Magnan, 2013). The increased motivational pull that game-like features bring also have a positive influence on learning (Young, et al., 2012). Recent research has suggested that learners consistently report higher levels of enjoyment and engagement when comparing standard tutoring systems and game-based systems, with similar performance outcomes, and GBT promoted more thorough responses on measures of learning after longer periods of time had passed (Jackson, Dempsey, & McNamara, 2011). When executed effectively, game-like design and instructional development can lead to optimal learning conditions in simulation-based training systems that adopt these features (Proske, Roscoe, & McNamara, 2014). This approach is immensely beneficial when considering current trends in training development that have led to a shift from the traditional classroom to stand-alone, individual training, often times without instructor oversight.

METHOD

Participants

Ten U.S. Soldiers participated in this experiment with one last-minute withdrawal due to a family emergency. Each participant was a recent graduate of the Maneuver Captain's Career Course (MCCC) located at Fort Benning, GA. Participants were randomly chosen from the course with an average (N = 10, M = 29.1, SD = 2.9) and years in service (N = 10, M = 5.7, SD = 2.2) representative of the course's population.

Research Objective

The purpose of this research was to examine the training effectiveness of the LG-RAID simulation in an operationally relevant environment. The method employed was an empirical assessment of the degree of training transfer of the LG-RAID condition in comparison to a control treatment. We evaluated the simulation using Kirkpatrick's model for evaluating training programs, especially focused on Level II (Learning Criteria), which is whether or not there was an increase or decrease in the student's knowledge or capability as a consequence of using the simulation. For this study, training effectiveness referred to whether or not there was a qualitative difference discovered in the tactical plans produced between treatment groups, where the tactical plan produced represented the participant's individual capability. This approach is congruent with Kirkpatrick's Level II assessment, allowing us to measure whether the simulation improved a trainee's capability or not, in comparison to the control treatment. Results of this study will also be used in the future to identify critical usability and functionality design recommendations for the simulation in order to optimize its future capabilities.

Design of Experiment

The experiment was conducted at the Maneuver Center of Excellence (MCoE) located at Fort Benning, GA using recent graduates of the MCCC as our sample. The experiment utilized one independent variable - training condition. Students were assigned a tactical mission by the MCCC Small Group Leader (SGL), in a classroom setting, and were given five hours to develop a tactically sound plan. Participants performed this tactical planning exercise either using the LG-RAID simulation software (experimental group) or by employing a more traditional method of planning, such as a mapboard (control group). Dependent variables included performance and survey responses. Performance was measured by the SGL's evaluation and assessment of the student's tactical plan, using the SGL Evaluation Survey, which we developed for this experiment. The SGL Evaluation Survey was composed of 11 distinct performance measures, each of which were scored on a 7-point Likert scale. Survey responses included the individual's completion of both the System Usability Questionnaire (Brooke, 1996) and Trainee Feedback Survey.

This experiment was conducted over the period of two days. On the first day, all participants completed their consent forms and demographics survey prior to training. Participants were then provided a block of instruction on the functionality of the LG-RAID simulation which included a brief PowerPoint overview of the simulation followed by a live demonstration of the simulation's capabilities. Participants were then provided guidance and practical exercises on how to create, run and modify various scenarios in the simulation (Figure 1). Prior to departing on day 1, participants were randomly assigned to either the LG-RAID or traditional training treatment for the following day's experiment.

Figure 1: Day 1 Training

The experiment was conducted on the second day of training. In both treatments, individuals were assigned a tactical mission by the SGL and directed to develop their best possible tactical plan within five hours. In the traditional treatment, individuals developed their tactical plan using legacy methods employed by the Army, such as mapboards and PowerPoint. In the LG-RAID treatment, individuals utilized the LG-RAID simulation software to rapidly visualize their initial plan, receive feedback, and iterate as necessary based upon the simulation's output (Figure 2). For both treatments, the student's objective was to create the best tactical plan that they were capable of, given the tools associated with each treatment.

Figure 2: Day 2 Experimental (LG-RAID) Treatment

The tactical scenario presented to students was derived from the MCoE's MCCC existing Program of Instruction (POI) and involved a Stryker company, with combined arms enablers, attacking enemy forces in an urban environment. All content, with the exception of the LG-RAID familiarization class, was developed and delivered by certified U.S. Army instructors who possessed intimate knowledge of the correct exercise solution set and were considered domain experts. During planning development, the only difference between treatments was the presence (or lack thereof) of the LG-RAID simulation software.

The instructor formally evaluated the individual's plan upon completion of the exercise. Performance assessment was conducted in a single-blind manner. The student briefed the SGL on his plan and the instructor evaluated the student's plan in accordance with the 11 distinct performance measures contained in the SGL Evaluation Sheet developed for this experiment. Upon completion of the training exercise, participants completed two surveys: the System Usability Questionnaire (Brooke, 1996) and the Trainee Feedback Survey. In the next section, we discourse on the results of this experiment.

RESULTS

As previously discussed, student performance was assessed by an accredited U.S. Army SGL using the SGL Evaluation Survey, which we created for this experiment. The SGL Evaluation Survey was composed of 11 distinct performance measures, each of which was scored on a 7-point Likert scale. The 11 performance measures were then aggregated into five performance areas: Tactical Plan, Mission Command, Holistic Plan, Analysis and Total Performance. The Tactical Plan performance area examined whether the student's plan adhered to proper employment of tactics, techniques and procedures. Mission Command examined the student's exercise of authority and direction as the commander using mission orders to enable disciplined initiative within the commander's intent. The Holistic Plan performance area focused on whether the student's plan utilized all available resources, was complete, and included appropriate contingency plans. The Analysis performance area examined whether the student's plan addressed tactical risk and incorporated proper terrain and enemy analyses. Total Performance was the student's aggregated score of the four performance areas. Performance results of the two treatments are depicted in Table 1 and Figure 3.

Table 1: Performance Results

Treatment	LG-RAID		Traditional	
Performance Area	Mean	SD	Mean	SD
Tactical Plan	5.27	0.72	5.42	1.10
Mission Command	4.40	0.55	4.13	0.48
Holistic Plan	4.67	0.41	5.00	1.22
Analysis	5.20	1.17	4.75	0.88
Total Performance	**4.93**	**0.63**	**4.89**	**0.79**

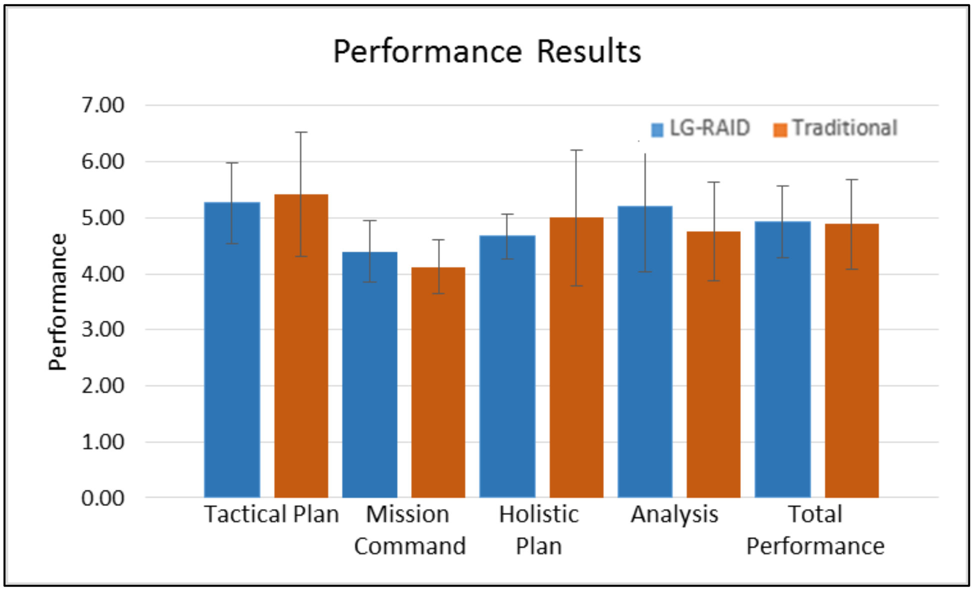

Figure 3: Performance Results

For our quantitative analysis, a series of Student's t-Tests were performed to examine whether there were significant differences in performance between treatments. Since the sample size was small, these results should be treated as preliminary. Future data collection events will be conducted to increase the sample size and thus reduce the probability of committing a Type II error, whereby we mistakenly fail to reject the null hypothesis (no difference in performance between treatments) when in fact the null was not true. Statistical analysis found there was no significant effect of treatment on Tactical Plan performance $t(5) = -0.23$, $p = 0.82$, 95% CI [-1.8, 1.5], Mission Command performance $t(7) = 0.8$, $p = 0.45$, 95% CI [-0.5, 1.1], Holistic Plan performance $t(4) = -0.53$, $p = 0.63$, 95% CI [-2.2, 1.5], Analysis performance $t(7) = 0.66$, $p = 0.53$, 95% CI [-1.2, 2.1], nor Total performance $t(6) = 0.08$, $p = 0.94$, 95% CI [-1.2, 1.2]. In recognition of the small sample size, a non-parametric Mann-Whitney test also indicated no difference in the average performance for the LG-RAID group (*Mdn* = 3.0) and the control group (*Mdn* = 6.5), $U = 19$, $p = .19$.

Qualitative analysis was collected through a post-hoc guided discussion between the Principal Investigator (PI) and SGL. The students' treatments were revealed to the SGL (post-hoc) in order that he may detect and describe general trends, differences and similarities between the two groups' performance. The SGL noted that students in the LG-RAID treatment did not employ phased operational planning but were better able to articulate their plan's sequence of events. Similarly, the SGL observed that LG-RAID students produced less structured orders than their control group peers. The SGL also noted that while the control group adhered more closely to the MCCC's operations order structure, the LG-RAID treatment's plans contained more detail. Finally, while these observations are valuable, it should be noted that the SGL's holistic assessment was congruent with our quantitative analysis; there was no significant difference in performance between both treatments.

DISCUSSION

Conclusion

While the results described above found no significant statistical difference in performance, we caution that the low sample size renders our results as preliminary. Future data collection events will be executed so as to mitigate the low statistical power attributed to the low density of available MCCC graduates. Nevertheless, the results described above provide initial evidence to support the notion that lightweight game-based constructive simulation is an effective training simulation tool.

In concert with the MCoE's MCCC, we randomly assigned the most qualified tactical planners in the Army to either the LG-RAID condition or control treatment and then compared their performance in the development of a tactical plan. With the MCCC, we developed the SGL Evaluation Survey which identified 11 distinct performance measures that compose the totality of a tactical plan. We then aggregated those results into four performance areas: Tactical Plan, Mission Command, Holistic Plan and Analysis. Total Performance was calculated by averaging the student's performance in the four respective performance areas. For this study, we defined training effectiveness as whether or not a student's capability improved as a result of using the LG-RAID simulation, which adheres to level II of Kirkpatrick's model for evaluating training programs.

We found no significant difference in performance between both treatments, despite two important factors. First, the LG-RAID treatment incurred an approximate 20% time penalty due to the requirement that the student briefing be conducted in the control group's format. We expand upon this in the Lessons Learned section below. Secondly, it is noteworthy to highlight that the students received only an abbreviated exposure to the simulation prior to employing the software in an operational setting. In light of the above, we believe that lightweight game-based constructive simulation shows promise as an effective training simulation tool.

The LG-RAID simulation primarily employed three gaming elements within the simulation to improve Soldier performance in tactical planning. These elements included: challenge, uncertainty and progress. The first gaming element, challenge, refers to a problem set that is difficult to solve. In the case of this experiment, Soldiers were assigned a tactical mission to destroy a credible, free-thinking adversary while adhering to U.S. Army doctrine. Challenge was created through the use of the game's validated, simulated enemy force that required soldiers to think and plan as if the mission was real. This element promoted critical analysis and deep thinking by Soldiers, which served as a desired learning outcome for the MCCC. Uncertainty, the second employed gaming element, denotes that the outcome of the simulation is not known in advance. LG-RAID employs novel game theory which supports, to a certain extent, non-deterministic and stochastic simulated training outcomes. Thus, like any live enemy force encountered, a Soldier's particular simulated engagement may have taken multiple paths, none of which could be predicted, with certainty, in advance. Uncertainty required Soldiers to develop plans that adhered to Army doctrine, which has continually evolved in response to changing threats the Army encounters. This served as another MCCC desired learning outcome. The final gaming element was progress and refers to the simulation's ability to depict and playback the individual's performance. Progress was depicted primarily through the simulation's statistical output, which displayed, in both real-time and post-hoc, both friendly and enemy casualty information, by individual soldier and vehicle platform. Progress provided feedback that allowed Soldiers to refine their plan, generate a new estimate and then re-execute the simulation with the calibrated plan. In this manner, Soldiers could generate the most optimal qualitative tactical plan, which also serves as the desired learning outcome for the MCCC.

Lessons Learned

This experiment placed the LG-RAID treatment at a time disadvantage. In order for the study to maintain the integrity of the single blind format, it was necessary to require students in the experimental treatment to transpose their final plan from the LG-RAID software to a traditional map overlay (the control group's format). While this maintained the single-blind design, it reduced available planning time to this treatment by approximately 20%. In order to not place the LG-RAID treatment at a time disadvantage in the future, we will need to develop a third briefing modality whereby both treatments are forced to convert their plan into a neutral format.

Another lesson learned involved the amount of time spent on initial training with the LG-RAID simulation. An anecdotal observation, confirmed through informal after-action review discussions with Soldiers in the experimental treatment, revealed that these Soldiers felt they would have performed better on their evaluated plan if they had a better familiarization with the simulation. Proficiency is a function of well-structured training and the appropriate amount of time to execute that training. In the future, we are considering adding a day of simulation training to determine if performance improves accordingly.

Recommendations for Future Research

As previously discussed, future data collection events are planned with the MCoE's MCCC and represent the primary future direction of research for this effort. The low density of qualified Soldiers, coupled with the demanding tempo of the MCCC Program of Instruction result in the need for multiple data collection events instead of a singular data collection approach. Additionally, future research will examine the usability and functionality feedback obtained at this, and future, data collection events. This feedback will guide future software development activities as well as interface improvements, which will be discoursed in our future work. Finally, the two topics discoursed in Lessons Learned, namely a neutral briefing format that does not penalize just one treatment and the addition of training time for simulation familiarization will be incorporated in future design of experiment adjustments.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to both the Maneuver Center of Excellence's Maneuver Captains Career Course cadre as well as the course graduates for their support of this and future experiments. These officers represent the best tactical planners in the world.

REFERENCES

Belanich, J., Orvis, K. L., & Sibley, D. (2013). PC-based game features that influence instruction and learner motivation. *Military Psychology, 25*(3), 206-217.

Bergeron, B. (2006). *Developing Serious Games (Game Development Series).* Charles River Media.

Brooke, J. (1996). A "quick and dirty" usability scale. In P. Jordan, B. Thomas, & B. Weerdneester, *Usability Evaluation in Industry* (pp. 189-194). London: Taylor & Francis.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*(2), 661-686. doi:10.1016/j.compedu.2012.03.004

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441-467. doi:10.1177/1046878102238607

Gega, L., Norman, I. J., & Marks, I. M. (2007). Computer-aided vs. tutor-delivered teaching of exposure therapy for phobia/panic: Randomized controlled trial with pre-registration nursing students. *International Journal of Nursing Studies, 44*(3), 397-405. doi:10.1016/j.ijnurstu.2006.02.009

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious game as new educational tools: how effetive are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29*(3), 207-219. doi:10.1111/j.1365-2729.2012.00489.x

Hodson, D. D., & Hill, R. R. (2013). The Art and Science of Live, Virtual and Constructive Simulation for Test and Analysis. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*.

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.

Jackson, G. T., Dempsey, K. B., & McNamara, D. S. (2011). Short and Long Term Benefits of Enjoyment and Leaning within a Serious Game. *Artificial Intelligence in Education: Proceedings of the 15th Internation Conference on Artificial Intelligence in Education,* 139-146. doi:10.1007/978-3-642-21869-9_20

Keebler, J. R., Jentsch, F., & Schuster, D. (2014). The effects of video game experience and active stereoscopy on performance in combat identification tasks. *Human Factors, 56*(8), 1482-1496. doi:10.1177/0018720814535248

Orlansky, J., Dahlman, C., Hammon, C., Metzko, J., Taylor, H., & Youngblut, C. (1994). *The Value of Simulation for Training.* Alexandria: Institute for Defense Analyses.

Proske, A., Roscoe, R. D., & McNamara, D. S. (2014). Game-based practice versus traditional practice in computer-based writing strategy training: Effects on motivation and achievement. *Educational Technology Research And Development, 62*(5), 481-505. doi:10.1007/s11423-014-9349-2

Richardson, A. E., Powers, M. E., & Bousquet, L. G. (2011). Video game experience predicts virtual, but not real navigation performance. *Computers in Human Behavior, 52*(2). doi:10.1016/j.artmed.2011.04.003

Riecken, M., Powers, J., J. C., Numrich, S. K., Picucci, P. M., & & Kierzewski, M. (2013). The Value of Simulation in Army Training. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).* National Training Systems Association.

Roman, P., & Brown, D. (2008). Games – Just How Serious Are They? *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (pp. Vol. 2008, No. 1). NTSA.

Serge, S. R. (2014). A Multimedia Approach to Game-Based Training: Exploring the Effects of the Modality and Temporal Contiguity Principles on Learning in Virtual Environment. *[Unpublished Dissertation]*.

Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The Effects of Static and Adaptive Performance Feedback in Game-Based Training. *Computers in Human Behavior, 29*(3), 1150-1158. doi:10.1016/j.chb.2012.10.007

Sitzmann, T. (2011). A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. *Peronnel Psychology, 64*(2), 489-528. doi:10.1111/j.1744-6570.2011.01190.x

Sotomayor, T., & Proctor, M. (2009). Assessing Combat Medic Knowledge and Transfer Effects Resulting from Alternative Training Treatments. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 121-134.

Stevens, J., Eifert, L., Reed, D., Diaz, E., & Umanskiy, O. (2014). Lessons Learned in Creating an Autonomous Driver for OneSAF. *I/ITSEC* (pp. 2661-2672). Orlando, FL: NTSA.

Stilman, B. (2000). Linguistic Geometry: From Search to Construction. Kluwer (now Springer).

Stilman, B., Yakhnis, V., & Umanskiy, O. (2007). Strategies in Large Scale Problems. In Ed. by A. Kott (DARPA) and W. McEneaney (UC-San Diego), *Adversarial Reasoning: Computational Approaches to Reading the Opponent's Mind,* (pp. Chapter 3.3, pp. 251-285). Chapman & Hall/CRC.

Stilman, B., Yakhnis, V., & Umanskiy, O. (2010). Linguistic Geometry: The Age of Maturity. *Journal of Advanced Computer Intelligence and Intelligent Informatics*, Vol. 14, No. 6, pp. 684-699.

Weiner, G. M., Menghini, K., Zaichkin, J., Caid, A. E., Jacoby, C. J., & Simon, W. M. (2011). Self-directed versus traditional classroom training ofr neonatal resuscitation. *Pediatrics, 127*(4), 7133-719. doi:10.1542/peds.2010-2829

Whitney, S., Tempby, P., & Stephens, A. (2014). A Review of the Effectiveness of Game-based Training for Dismounted Soldiers. *Journal of defenese Modeling and Simulation,* 319-328.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., . . . Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research, 82*(1), 61-89. doi:10.3102/0034654312436980