

## Training Effectiveness Evaluation of Lightweight Game-based Constructive Simulation

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

**Ms. Latika Eifert**  
**Dr. Sean Mondesire**  
Army Research Laboratory (ARL)  
Human Research and Engineering Directorate  
Orlando, FL  
latika.eifert@us.army.mil  
sean@cs.ucf.edu

### ABSTRACT

This research is a continuation and expansion of our previous analysis of the efficacy of lightweight, constructive, game-based simulation. While the U.S. Army continues to employ constructive and game-based simulation for training, minimal research has been conducted on the training efficacy of these classes of simulation. 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 remained unchanged; to examine the training effectiveness of the U.S. Army's LG-RAID simulation in an operationally relevant environment. Qualified Soldiers were randomly assigned to one of two training treatments (LG-RAID simulation 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 on multiple occasions at Fort Benning, GA and performance was evaluated by accredited U.S. Army instructors. Results of this study continue to indicate that the LG-RAID simulation may be an effective training simulation.

### 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) and Mathematics Professor at Valencia College, 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.

**Dr. Sean C. Mondesire** is a Postdoctoral Research Fellow at the U.S. Army Research Lab (ARL) and holds a doctoral degree in computer science from the University of Central Florida. His research focus is on the expansion of virtual world technologies to be used in tactical, military training.

# Training Effectiveness Evaluation of Lightweight Game-based Constructive Simulation

**Dr. Jonathan Stevens**  
**Dr. Stephen R. Serge**  
 University of Central Florida (UCF)  
 Orlando, FL  
[jonathan.stevens@knights.ucf.edu](mailto:jonathan.stevens@knights.ucf.edu)  
[sserge@ist.ucf.edu](mailto:sserge@ist.ucf.edu)

**Ms. Latika Eifert**  
**Dr. Sean Mondesire**  
 Army Research Laboratory (ARL)  
 Human Research and Engineering Directorate  
 Orlando, FL  
[latika.eifert@us.army.mil](mailto:latika.eifert@us.army.mil)  
[sean@cs.ucf.edu](mailto:sean@cs.ucf.edu)

## INTRODUCTION

The United States Army is placing more emphasis on training for major combat operations (MCO) as doctrinal changes stress leader development (Odierno, 2012). New approaches, as well as new technologies, are being evaluated for the Army's challenging transition back to MCO training (Eifert, Stevens, Reed, Umanskiy, & Stilman, 2015). One of these approaches is the increased use of advanced simulation-based training (SBT) to facilitate strategic and tactical training for military personnel. Specifically, the U.S. Army Training and Doctrine Command (TRADOC) has mandated that by 2020, approved SBTs must be rapid, responsive, and real-time to support future immersive and customizable soldier training.

SBT can be organized into three classes: live, virtual, and constructive simulations (Hodson & Hill, 2013). In *live* simulations, real people operate real systems in real, physical environments. By contrast, *virtual* simulations are comprised of real people operating synthetic systems in synthetic environments while *constructive* simulations contain synthetic entities operating synthetic systems in synthetic environments. Although live, virtual, and constructive are considered the major classes of simulation, a fourth classification, *gaming*, has emerged in the last ten years (Roman & Brown, 2008). In game-based simulation, real people train and interact with computer-based systems and environments (Bergeron, 2006). All four of these classifications continue to be deployed to suit the needs of diverse and expansive simulated training domains. In modern SBT, there are instances when the live, virtual, constructive, and gaming classifications overlap; this phenomenon is referred to as *blended training*. Blended training occurs when multiple simulation classes are present in a single training exercise to leverage the advantages of the overlapped types of simulation.

This research is a continued effort to evaluate the training effectiveness of the blended Linguistic Geometry Real-time Adversarial Intelligence and Decision-making (LG-RAID) simulation tool. LG-RAID is a light-weight software tool that is used to automate the process of generating intelligent, predictive and tactically-correct Courses of Action (COAs) for constructive exercise participants at the company echelon and below. The methodology behind the tool is based on Linguistic Geometry (LG), a theory derived from both Artificial Intelligence (AI) and Game Theory. The theory has demonstrated an ability to solve large-scale problems in near real-time (Stilman B. , 2014; Stilman, Yakhnis, & Umanskiy, 2010) and it is particularly useful in solving strategy-like problems, such as mission planning. Development efforts of the LG-RAID tool have focused on creating an easy-to-use tool that will allow the student-users to visualize their COAs in an instructor-facilitated class environment with little to no overhead. Low overhead simulation has been defined by the U.S. Army as able to operate on standard computers, learnable in less than four hours and requires no external adjudication (Allen & Sterrett, 2012).

Since initial development, LG-RAID's capabilities have matured. The next milestone in the tool's life-cycle is to transition it into an Army Program of Record (PoR) so Soldiers can leverage the training advantages of the technology. To accomplish this milestone, the U.S. Army Research Laboratory (ARL) and the University of Central Florida (UCF) Institute for Simulation and Training (IST) have collaborated on a multi-phased investigation to determine the training effectiveness and usability of LG-RAID for Soldier training. This work presents the results of the second phase of analysis, where the primary objective was to empirically assesses the training effectiveness of

LG-RAID simulation using a qualified population of Soldiers. The study's secondary objective of assessing the usability and functionality of the simulation will be published at a future time.

## **BACKGROUND**

### **LG RAID Background**

The current version and underlying principles of LG-RAID are based on a theory of Abstract Board Games in AI 1972 (Stilman B. , 2000; Stilman, Yakhnis, & Umanskiy, 2010). This approach is scalable and allows for the generation of simulated strategies of real world systems in near real-time. This outcome is attainable due to the use of an approach that utilizes a synthesis-based algorithm (i.e., construction of strategies within the Abstract Board) rather than a tree-based search algorithm. This, in turn, leads to an avoidance of the potentially exponential growth of probabilities to which tree-based approaches are vulnerable.

LG-RAID development was initiated by the Defense Advanced Research Projects Agency (DARPA) RAID program in 2004. The goal was to develop a decision support aid for lower tactical echelons (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. These same capabilities are now being leveraged to provide a lightweight capability for simulation-based training.

Within a military classroom environment, students are currently able to use LG-RAID to create, visualize, and test COAs based on presented mission parameters. Since LG-RAID is accessible via a standard web-browser, students may use a standard personal computer to interact with the simulation. Users can input mission data, such as friendly and enemy intelligence (i.e., units, equipment, specialized weaponry, etc.), alter specific terrain restrictions (i.e., No-Go zones), and develop their scheme of maneuver through the available execution matrix GUI. Based on these inputs, LG-RAID executes the scenario and provides tactically accurate and relevant results for review with very little additional input required from the user. The user is then provided an opportunity to review the outcome of their simulated COA and make adjustments to compare outcomes of different strategies. This approach is intended to provide a higher level of interaction from the student, as well as more opportunities to review and understand detailed tactical planning and increasing proficiency.

### **Game-Based Training (GBT)**

Game-derived characteristics (i.e., scores and points, earning trophies/rewards, avatar personalization, etc.) have become increasingly incorporated into computer-based simulation training systems (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). The ability for games to increase the motivational or rewarding characteristics of training has led to an increased interest in utilizing these feature in different GBT platforms (Belanich, Orvis, & Sibley, 2013; Proske, Roscoe, & McNamara, 2014).

Current simulation and game-based training research has often reported that GBT systems are a highly effective approach to training (Grossman, Heyne, & Salas, 2015). In some instances, research suggests similar or improved performance when compared to more traditional styles of training, such as classroom-based lectures (Proske, Roscoe, & McNamara, 2014; Gega, Norman, & Marks, 2007; Jackson & McNamara, 2013). Additionally, research has also reported that the reinforcement of training material and the ability to practice in a realistic simulated or gaming environment is beneficial when training material is well presented (Serge, Priest, Durlach, & Johnson, 2013; Weiner, et al., 2011). When comparing results between traditional methods of training and GBT, results often reveal that game-specific features are capable of promoting high levels of effective learning in simulated training systems (Proske, Roscoe, & McNamara, 2014; Sitzmann, 2011).

The game-like features of GBT also increase trainee engagement and motivation, which leads to increased involvement with the system when compared to other non-game types of training (Garris, Ahlers, & Driskell, 2002; Girard, Ecalte, & Magnan, 2013). This increased level of motivation also increases the likelihood of positive learning outcomes (Young, et al., 2012). GBT, and game-like qualities within these systems, can significantly increase the overall enjoyment, engagement, and learning outcomes from the training experience if executed appropriately.

## **METHOD**

### **Participants**

Twenty-two U.S. Soldiers participated in this experiment. Each participant was a recent graduate of the Maneuver Captain's Career Course (MCCC) located at Fort Benning, GA. Participants were randomly selected from a pool of recent course graduates and had an average age ( $N = 22$ ,  $M = 28.9$ ,  $SD = 2.3$ ) and years in service ( $N = 22$ ,  $M = 5.7$ ,  $SD = 1.9$ ) representative of the course's population.

### **Research Objective**

The purpose of this research was to investigate the training effectiveness of the LG-RAID simulation in an operationally relevant environment. This study's focus 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's efficacy using Level II of Kirkpatrick's model for evaluating training programs. Level II, Learning Criteria, is defined as 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 student's produced tactical plan represented his individual capability. This approach is consistent with Kirkpatrick's Level II assessment, allowing us to measure whether the simulation (experimental treatment) improved a trainee's capability or not, in comparison to the control treatment. Level I data and feedback (Reaction Criteria) were also collected and will be used in the future to identify critical usability and functionality design recommendations for the simulation, with those results to be published separately.

### **Design of Experiment**

The experiment was composed of a series of data collection events at the Maneuver Center of Excellence (MCoE) located at Fort Benning, GA. The experiment used recent graduates of the MCCC as our sample. Both the sample and population are a very low density specialty in the Army, therefore we executed multiple data collection events to maximize our sample size. 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 in one of two training conditions: either using the LG-RAID simulation software (experimental group) or by employing a more traditional method of planning taught at the MCCC, such as the use of a mapboard (control group). Dependent variables included performance and survey responses. Performance was objectively measured via a single-blind assessment by a designated SGL who had no knowledge of the experiment's design. The SGL evaluated each student's tactical plan using the SGL Evaluation Survey, which we developed for this experiment in conjunction with the MCoE. 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.

Each data collection event was two days in duration. On the first day, 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. Training consisted of a concise PowerPoint overview of the simulation followed by a live demonstration of the simulation's capabilities. Participants then engaged in active learning on the simulation for the rest of the first day. The group executed numerous practical exercises on how to create, run and modify various scenarios in the simulation (Figure 1). After the first day's training was completed, participants were randomly assigned to either the LG-RAID or traditional training condition by the Principal Investigator for the following day's actual experiment.



Figure 1: Day 1 Training

On the second day of data collection, all participants received an assigned tactical mission through an Operations Order brief from a MCCC SGL. From there, the two treatment groups separated and commenced their planning process in two distinct locations. In both treatments, individuals were directed to develop the best possible tactical plan within the allotted five hours. In the control or traditional treatment, individuals developed their tactical plan using legacy methods and tools employed by the Army, and taught at the MCCC, such as acetated mapboards and PowerPoint. In the LG-RAID treatment, individuals utilized the LG-RAID simulation software to rapidly develop and 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 qualitative tactical plan, utilizing the tools associated with each treatment.



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

The tactical scenario presented to students was consistent amongst each data collection event and was derived from the MCoE's MCCC existing Program of Instruction (POI). The scenario consisted of a U.S. Army Stryker company, augmented 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 Small Group Leaders of the MCCC who possessed detailed knowledge of the training exercise and were considered domain experts. The significant difference between treatments (training conditions) was the presence (or lack thereof) of the LG-RAID simulation software.

A SGL with no knowledge of the design of experiment formally evaluated each individual's plan upon completion of the exercise. Performance assessment was conducted in a single-blind manner for study validity as well as to mitigate challenges that SGLs face in measuring student performance (Leibrecht, Tucker, Haverty, Blankenbeckler, & Green, 2009). Each student briefed the SGL on his plan, in the control group's analog format, and the SGL evaluated each 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. Performance results are discussed in the next section.

## RESULTS

Student performance was assessed, post-hoc, by an accredited U.S. Army SGL using the SGL Evaluation Survey, which we created with the MCCC cadre 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 U.S. Army 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 and comparisons between the two treatments are displayed in Table 1 and Figure 3.

Table 1: Performance Results

| Treatment         | LG-RAID |      | Traditional |      |
|-------------------|---------|------|-------------|------|
| Performance Area  | Mean    | SD   | Mean        | SD   |
| Tactical Plan     | 4.45    | 1.66 | 4.91        | 1.13 |
| Mission Command   | 3.82    | 1.50 | 4.27        | 0.61 |
| Holistic Plan     | 4.09    | 1.32 | 4.67        | 1.00 |
| Analysis          | 4.73    | 1.37 | 4.61        | 1.05 |
| Total Performance | 4.31    | 1.39 | 4.64        | 0.88 |

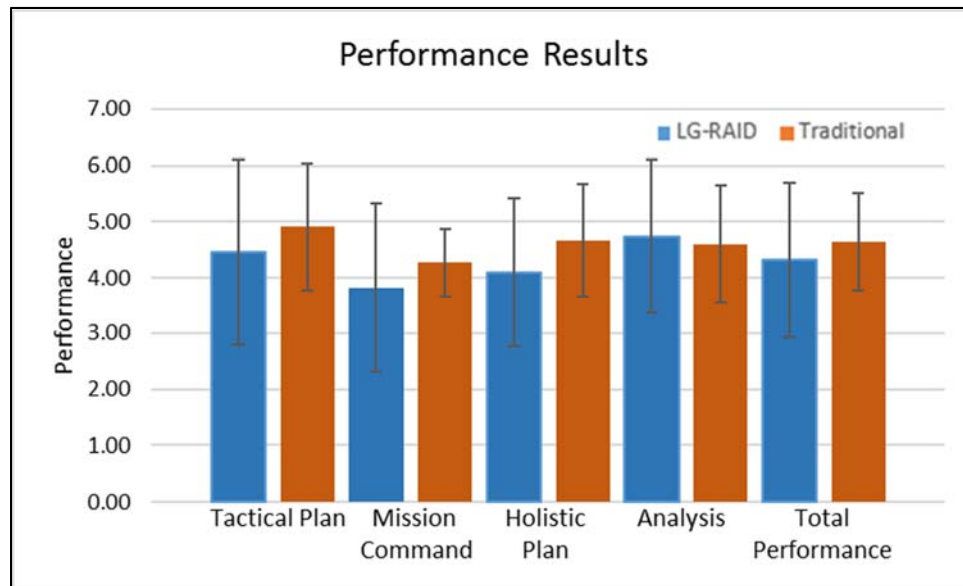


Figure 3: Performance Results

For our statistical analysis, we again employed a series of Student's t-Tests to examine whether there were significant differences in performance between treatments. While the sample size remained small, our results are strengthened from our previous work due to additional data collection events conducted. It should be noted that the low density of this particular population of soldiers makes a significantly larger sample size virtually impossible to achieve. Our increased sample size did reduce our probability of committing a Type II error, whereby we mistakenly fail to reject the null hypothesis (concluding 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  $t(18) = -0.75, p = 0.46, 95\% \text{ CI } [-1.73, 0.82]$ , Mission Command performance  $t(13) = -0.93, p = 0.37, 95\% \text{ CI } [-1.51, 0.6]$ , Holistic Plan performance  $t(19) = -1.16, p = 0.26, 95\% \text{ CI } [-1.62, 0.47]$ , Analysis performance  $t(19) = 0.23, p = 0.82, 95\% \text{ CI } [-0.97, 1.21]$ , nor Total performance  $t(17) = -0.67, p = 0.51, 95\% \text{ CI } [-1.37, 0.71]$ . In further recognition of the small sample size, a non-parametric Mann-Whitney test also indicated no significant difference in the average performance between the LG-RAID group ( $Mdn = 13.0$ ) and the control group ( $Mdn = 11.0$ ),  $U = 131, p = .80$ .

Qualitative analysis was collected through a post-hoc assessment tool completed by the SGL for every student, with trend analysis conducted by the Principal Investigator (PI) and research team. In this manner, general trends, differences and similarities between the two groups' performance could be isolated and analyzed. The SGL assessment revealed several noteworthy observations and potential trends. First, students who were placed in the experimental treatment (LG-RAID) used less clearly defined phased planning than their control group peers, which was a congruent finding from our pilot study. This may be a coping technique by students as they create their plan to align with the planning template offered by the simulation. However, the SGL's observations indicate that while the experimental treatment group's plans were less structured, they contained more detail. This observation may be a result of the rich detail the simulation provides to the student. Another trend discovered from the SGL's post-hoc qualitative assessment indicated that the LG-RAID users did not employ indirect fires as effectively as their control group peers. This is concerning since maneuver commanders are trained at the MCCC to be combined arms commanders, not just maneuver commanders, and thus should incorporate all critical enablers into their plan. This observation may be a direct result of the time penalty incurred by the experimental treatment group, who were forced to convert their plan to the control group's briefing format, so as to maintain the single-blind format of the study. By doing this, it appears that indirect fire planning may have been the area that students did not have adequate time to plan for. Finally, while these trends provide extremely valuable feedback, it should be noted that the SGL's qualitative, post-hoc assessment was consistent with the study's quantitative analysis; there was no significant difference in performance found between both treatments.



## DISCUSSION

### Conclusion

In this study we found no significant difference in Soldier performance between the experimental treatment and the control treatment. We believe, however, that we have discovered evidence that lightweight game-based constructive simulation is an effective training simulation tool. There were a number of design of experiment constraints placed on the experimental condition group that will be discoursed later that support our position. While our initial sample size from the pilot study was small, the focus of this paper was to expand our pilot study findings to include the results of additional data collection events conducted at the MCCC.

In cooperation with the MCoE's MCCC, we assessed the training effectiveness of the LG-RAID simulation at level II of Kirkpatrick's model for evaluating training programs. We randomly assigned recent graduates of the MCoE's MCCC to either the LG-RAID condition or control treatment and then compared their tactical planning performance using a training scenario developed by MCCC cadre. Soldier performance, defined as the tactical plan produced by the Soldier, was evaluated by an accredited SGL in a single-blind manner, using a rubric developed specifically for this experiment. The SGL Evaluation Survey was composed of 11 distinct performance measures that constitute all the major elements of a complete tactical plan. Those elements were then totaled into four performance areas: Tactical Plan, Mission Command, Holistic Plan and Analysis and finally averaged into a Total Performance category. For this study, we treated training effectiveness as whether or not a student's capability improved as a result of using the LG-RAID simulation.

Although we found no statistically significant difference in performance between the two treatments, we do believe that lightweight game-based constructive simulation is an effective training simulation tool. There were two major, yet inadvertent, design of experiment constraints placed on soldiers in the experimental treatment that we believe justify this assertion. First, soldiers in the LG-RAID training condition were required to manually transform their final plan, in digital format, to the control treatment's analog format. This requirement was placed upon the experimental group so as to maintain the integrity of the study's single-blind format. However, this requirement also reduced the experimental treatment's planning time by approximately 15% in comparison to the control group, who had no such conversion requirement. Despite the reduced planning time, Soldiers using the LG-RAID simulation were able to overcome this temporal deficit and achieve qualitatively similar results to the control group. Secondly, soldiers in the experimental treatment only received a one-day, abbreviated training session on the simulation and its capabilities. A single day of training is insufficient to make one an expert in the simulation, especially when factoring in the complexity of the LG-RAID simulation. Despite this abbreviated training session, soldiers in the experimental treatment were able to utilize the limited training they did receive and achieve qualitative parity with respect to their control group counterparts.

The LG-RAID simulation served as a lightweight, constructive simulation by providing soldiers with the ability to rapidly develop and execute automated simulation runs and then presenting the advantages and disadvantages of each run through both numerical statistical output as well as intuitive, animated playback of the missions. The simulation provided soldiers with impressive augmented planning capabilities in a very low-overhead fashion. There were minimal personnel and no special computational resources needed to leverage the simulation's features, demonstrating how classroom-based military education can employ constructive simulation efficiently and effectively to achieve designated learning outcomes.

The LG-RAID simulation also served as a game-based simulation as it employed two core gaming elements (challenge and uncertainty) to improve soldier tactical planning performance. Challenge was created by presenting soldiers with a difficult problem set; to destroy a credible, free-thinking adversary while adhering to U.S. Army doctrine. The simulation utilizes a validated, simulated enemy force that acts and reacts in a highly intelligent manner. This forces exercise participants to critically analyze the situation, which is a desired learning outcome for the MCCC. Uncertainty is created through the simulation's use of non-deterministic and stochastic outcomes generated by the underlying algorithms driving the simulation. In this manner, no particular soldier's scenario outcome(s) could be predicted, with certainty, in advance. Soldiers must address this uncertainty by developing plans that adhere to U.S. Army doctrine, which is another MCCC desired learning outcome. Doctrine is an ever-evolving knowledge set that responds to changing threats the Army encounters. Tactical planners must use doctrine as the foundation for their plan; this is the crux of the 16-week MCCC.

## Recommendations for Future Research

Future research will be conducted on the training effectiveness of the LG-RAID simulation using the United States Marine Corps USMC) as both our sample and population. In this manner, we will be able to determine and measure the training effectiveness of the simulation with a similar, yet different, population. As a direct result of our current research, the USMC has approached us to conduct a similar study for their service in order to determine whether lightweight, constructive game-based simulation is a potential training solution for their current and future needs. Additionally, future research will examine the usability and functionality feedback obtained at these data collection events. This feedback will guide future software development and interface improvement activities. Finally, experimental design adjustments, 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 data collection events.

## ACKNOWLEDGEMENTS

The authors wish to again 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.

## REFERENCES

- Allen, C., & Sterrett, J. (2012, January). A Battle in Every Classroom: Using Low-Overhead Simulations to Produce Experienced Captains and Majors. *Infantry*, pp. 44-47.
- 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
- Eifert, L., Stevens, J., Reed, D., Umanskiy, O., & Stilman, B. (2015). Adaptive Simulation-Based Training for a Complex World. *The International Workshop on Applied Modeling and Simulation* (pp. 1-7). Bergeggi, Italy: WAMS 2015: ISBN 978-88-97999-18-8; Bruzzone, Fadda, Fancello, Piera Eds.
- 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 effective 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 Learning within a Serious Game. *Artificial Intelligence in Education: Proceedings of the 15th International 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
- Leibrecht, B., Tucker, J., Haverty, R., Blankenbeckler, P., & Green, D. (2009). *Metrics for Assessing Cognitive Skills in the Maneuver Captains Career Course*. Fort Benning, GA: United States Army Research Institute.
- Odierno, R. (2012). The US Army in a Time of Transition: Building a Flexible Force. *Foreign Affairs*, 91, 7.
- 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 (IITSEC)*. National Training Systems Association.
- Roman, P., & Brown, D. (2008). Games – Just How Serious Are They? *Interservice/Industry Training, Simulation, and Education Conference (IITSEC)* (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. *Personnel 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. *IITSEC* (pp. 2661-2672). Orlando, FL: NTSA.
- Stilman, B. (2000). *Linguistic Geometry: From Search to Construction*. Kluwer (now Springer).
- Stilman, B. (2014). What Is the Primary Language. In *Artificial Intelligence and Soft Computing* (pp. 558-569). Springer International Publishing.
- 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 of 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 Defense 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