Measuring Visual Displays' Effect on Novice Performance in Door Gunnery

Dr. Jonathan Stevens
Army Research Laboratory (ARL)
Orlando, FL
jonathan.a.stevens@outlook.com

Dr. Peter Kincaid
University of Central Florida (UCF)
Orlando, FL
pkincaid@ist.ucf.edu

ABSTRACT

The purpose of this paper is to present the results of our recent experimentation involving a novice population performing aerial door gunnery training in a mixed reality simulation. Specifically, we examined the effect that different visual displays had on novice soldier performance; qualified infantrymen with machine gun experience. The results of this study differed from the findings of our first study, which utilized an expert population of qualified helicopter crew members.

The U.S. Army continues to develop new and effective ways to use simulation for training. One example is the Non-Rated Crew Member Manned Module (NCM3), a simulator designed to train helicopter crewmembers in critical, high risk tasks. Novice participants were randomly assigned to one of two visual display treatments (flat screen or Head-Mounted Display) and executed three aerial door gunnery training scenarios in the NCM3. Independent variables were visual display, trial, immersive tendency and simulator sickness questionnaire scores. Dependent variables included performance, presence and simulator sickness change scores. The performance results of this study differed from our first study and indicated there was a main effect of visual display on performance. However, both visual treatment groups experienced the same degree of presence and simulator sickness. Results of this study indicate that higher immersive simulation may lead to better performance for a novice population.

ABOUT THE AUTHORS

Dr. Jonathan Stevens is a retiring Lieutenant Colonel in the United States Army and a Research Program Manager of the Simulation & Training Technology Center (STTC), a subordinate organization of the Research Development & Engineering Command's (RDECOM) Army Research Laboratory (ARL). LTC Stevens has over 22 years of military experience as an Infantry and Acquisition Corps officer.

Dr. Peter Kincaid is a Graduate Research Professor at the University of Central Florida (UCF) and the co-director of the Modeling and Simulation graduate program at UCF. Dr. Kincaid's areas of research include training systems analysis, instructional technology and human factors. He has 30 years applied research and university teaching experience.

Measuring Visual Displays' Effect on Novice Performance in Door Gunnery

Dr. Jonathan Stevens
Army Research Laboratory (ARL)
Orlando, FL
jonathan.a.stevens@outlook.com

Dr. Peter Kincaid
University of Central Florida (UCF)
Orlando, FL
pkincaid@ist.ucf.edu

INTRODUCTION

Continuation of Research

The purpose of this paper is to present the results of our recent experimentation involving a novice population performing aerial door gunnery training in a mixed reality simulation. Specifically, we examined the effect that different visual displays had on novice soldier performance; defined as qualified infantrymen with machine gun experience. The results of this study differed from the findings of our first study, which utilized an expert population of qualified helicopter crew members. This study is a continuation of the research described in our previous work (Stevens & Kincaid, 2014).

Motivation for Research

There is a strong belief in the United States Army, particularly in the aviation community (Stewart, Johnson, & Howse, 2008), that the greater the degree of realism in a virtual simulation, the more effective that simulation is. Similarly, there exists a strong bias that the newer the technology in the simulator is, the more effective that simulator must be (Schout, Hendrikx, Scheele, & Scherpbier, 2010). However, little scientific research exists that support these notions (Borgvall, 2013) (Scales, 2013). In fact, a recent Government Accountability Office (GAO) report indicated that the Army "has not established metrics or indicators to assist them in more broadly measuring the impact of simulation-based training on improving the performance or proficiency of servicemembers" (United States Government Accountability Office, 2013).

U.S. Army simulation acquisition design decisions are occasionally made not based on sound scientific evidence, but rather to satisfy the user community's wants (Stevens & Kincaid, 2014). One such design trade-off, described below, and not grounded in science, served as the basis for this research.

The Non-Rated Crew Member Manned Module (NCM3)

Crew members, crew chiefs and flight engineers of rotary wing aircraft are classified as non-rated crew members (NCMs) in Army aviation vernacular, meaning they cannot fly the aircraft. However, NCMs must perform critical tasks such as aerial door gunnery, sling load operations and crew coordination. Until very recently though, these tasks could only be trained in a live environment (Marton, 2008). However, live training is costly in terms of fuel, ammunition and aircraft maintenance as well as being considered a high-risk endeavor. Thus dangerous tasks, such as aerial door gunnery, are infrequently trained by crew members and crew chiefs. To exasperate the situation, in OPERATION IRAQI FREEDOM, the U.S. Army experienced a shortage of NCMs and thus had to employ infantry soldiers to serve as door gunners (Curran, 2003).

The Non-Rated Crew Member Manned Module (NCM3) was fielded by the U.S. Army in 2011 in order to train NCMs in those critical, high risk tasks. The NCM3 is a mobile, transportable, multi-station virtual simulation device that can replicate the back of either a UH-60 (Blackhawk) or CH-47 (Chinook) helicopter. The NCM3 is capable of training aerial door gunnery in one of two visual conditions: Head-Mounted Display (HMD) or LCD flat panel screen.

The dual visual display capability of the NCM3 exists as a result of a negotiated trade-off conducted between the government acquisition office and the user community during the simulation's design phase (Stevens & Kincaid, 2014). The original design did not call for the use of any LCD displays in the trainer. However, the user community was insistent that aerial door gunnery must be capable of being trained in a non-HMD mode. Thus a trade-off was negotiated whereby a less expensive HMD, with a narrower field of view, was procured and four LCD flat screen displays were installed to support aerial door gunnery training. This trade-off resulted in a training device that was now capable of supporting mixed reality aerial door gunnery training via HMD or LCD display.

This trade-off was conducted in a non-scientific manner, with no empirical knowledge of what the training effect would be and runs contrary to Department of Defense modeling and simulation best-practice investment strategies (Aegis Technologies, 2008). In this study, we examine the effect that these different visual displays have on performance of aerial door gunnery.

BACKGROUND

Milgram & Colquhoun's (1999) global taxonomy of mixed reality display integration (Figure 1) remains the taxonomy this study adheres to. The three dimensions of the taxonomy consist of the Real-Virtual continuum, Congruence and Centricity. This study compares the effect on performance between a traditional visual display and an immersive Head Mounted Display (HMD) operating in the red oval of the MR Global Taxonomy. Thus the NCM3, as employed in this study, is a mixed reality, egocentric, highly congruent training simulator.

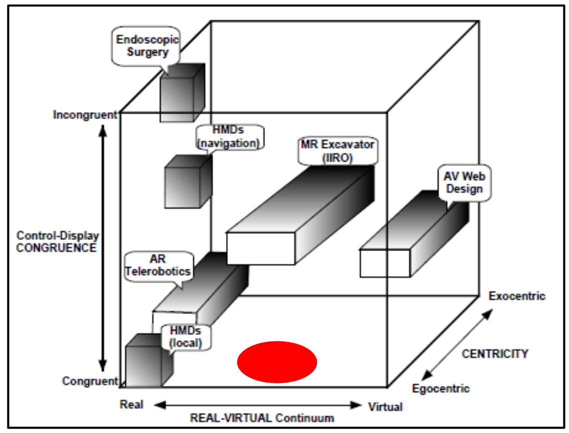

Figure 1. Global Taxonomy of MR Display Integration

The goal of a virtual simulation is generally to maximize the degree of skill transfer to the trainee, where transfer is the application of knowledge, skills and abilities acquired during training and applied to the environment they are normally used (Muchinsky, 2000), (Alexander, Brunye, Sidman, & Weil, 2005). Transfer has been empirically demonstrated to occur from a virtual to a real environment (Harrington, 2011) (Blow, 2012) (Seymour, et al., 2002) (Hays, Jacobs, Carolyn, & Salas, 1992) but this approach can be both difficult and costly. Thus a more common approach is to measure transfer, or the degree of learning, in the simulator itself. This study employs that methodology.

Fidelity is defined as the degree to which the virtual environment is indistinguishable from the real environment (Waller, Hunt, & Knapp, 1998). An effective training simulation need only possess a sufficient level of fidelity; the highest level of fidelity is not required for effective training (Summers, 2012) (Thorpe, 2010) (Borgvall, 2013) (Jentsch & Bowers, 1998). Immersion is primarily a function of the simulation's technology while presence refers to the trainee's subjective, personal experience in the simulation. Similar to fidelity, it is commonly believed that more immersion and greater presence induced by the simulator are correlated with higher transfer. However, this relationship is not clearly supported in the literature (McMahan, Bowman, Zielinski, & Brady, 2012) (Mikropoulos & Natsis, 2011) (Dalgarno & Lee, 2010) (Persky, et al., 2009) (Selverian & Sung, 2003).

Simulator sickness (SS) is defined as "the unwanted side effects and aftereffects that may result from using simulators, but does not result from similar use of the actual equipment" (Knerr, 2007). SS represents a major distraction in simulation based training due to lost training time and possible negative habit transfer. HMDs are becoming more common in military virtual simulators but have been identified as primary causes of SS.

Little research has been conducted examining the effect of different visual solutions in a virtual simulation. When experimentation has been conducted, most empirical results fail to prove that a training benefit exists when comparing the use of HMDs to more traditional visual displays (Barnett & Taylor, 2012) (Santos, et al., 2009) (Knerr, 2007) (Jacquet, 2002). Based on this, the generally higher cost of HMDs may not be justified due to the lack of a proven corresponding training benefit when compared to a lower cost visual display.

METHOD

Participants

The Novice group was composed of U.S. Army infantry soldiers with prior M240 machine gun experience. Soldiers were members of Bravo Company, 2nd Battalion, 124th Infantry Regiment, Florida Army National Guard. The sample size of 76 soldiers was 100% male and 0% female with a mean age of 26.2 (n = 76, M = 26.2, SD = 6.0).

Experimental Objectives

The four experimental objectives remained the same from our prior study, and are outlined below:

- The primary experimental objective (hypothesis 1) was to evaluate different visual displays' effect on performance in a Mixed Reality Aerial Door Gunnery Training environment.
- The second experimental objective (hypothesis 2) was to determine if different visual displays resulted in a difference in the level of simulator sickness experienced in a Mixed Reality Aerial Door Gunnery Training environment.
- The third experimental objective (hypothesis 3) was to determine if there exists a relationship between an individual's immersive tendency score and their performance and level of presence in a Mixed Reality Aerial Door Gunnery Training environment.
- The fourth sub-objective (hypothesis 4) was to examine if there exists a difference in the level of presence experienced when performing Mixed Reality Aerial Door Gunnery Training, based on visual display.

Apparatus

For this experiment, one NCM3 training device was transported to the unit's location for the company's drill weekend. The primary treatment of this experiment remained visual display. Participants either performed the task in stereoscopic HMD (NVIS nVisor MH60) or 46" flat panel screen condition. Additional key equipment included the use of the NCM3's demilitarized M240 Medium Machine Gun (MG) with simulated recoil, simulated flight helmet, emulated aperture and InterSense IS-900 tracking system.

Four questionnaires were employed for this second study. A demographic survey was used as the screening mechanism by the Principal Investigator to ensure that all participants had prior MG experience and were not qualified NCMs. The Simulator Sickness Questionnaire (SSQ) (Kennedy, Lane, Berbaum, & Lilienthal, 1993) was used in this study to measure simulator sickness (SS). The SSQ is a self-reporting symptom checklist that includes 16 symptoms associated with SS, rated on a four level scale. The symptom scores are then aggregated by subscale (Nausea, Oculomotor Discomfort, Disorientation) and Total Severity, with the score changes used to determine the impact of the simulation on participants' physiological state. The Immersive Tendencies Questionnaire (ITQ) (Witmer & Singer, 1998) was also utilized and is an 18 question (7 point scale) self-reporting checklist that measures participants' degree of potential for immersion. Participants' aggregated ITQ score was used as a predictor variable in this experiment. The Presence Questionnaire (PQ) (Witmer & Singer, 1998) was the fourth questionnaire employed and is a 24 question (7 point scale) self-reporting checklist that measured participants' degree of presence experienced during this experiment.

Four aerial door gunnery scenarios, created with the assistance of a subject matter expert (SME), were again employed for this second study. The familiarization scenario was developed in order to provide participants with the opportunity to familiarize themselves with the simulator, machine gun and assigned visual treatment. Three formal scenarios (first, second and third trial) were created for the conduct of the actual experiment. All scenarios emphasized kinetic engagements, with the participant's mission being to destroy as many enemy targets as possible with the aircraft's machine gun. Scenarios were approximately five minutes in duration and contained a maximum of 50 targets that could be destroyed by a soldier. Performance results, defined as the number of enemy targets destroyed, were captured at the I/O station.

Experimental Procedure

Prior to starting the experiment, in mass formation, participants were provided an overview brief by the Principal Investigator that covered topics such as the study's purpose, task/conditions/standards as well as a mission brief. In smaller groups, each participant then read and signed their consent forms and were subsequently randomly assigned to one of two visual display groups: HMDs or flat screen display. In groups of four, participants completed the three pre-test surveys and then executed the familiarization and three formal scenarios (trials one, two and three). All commands were given over the simulator's communication network from the I/O. Scoring was accomplished at the I/O station. Upon completion of the four scenarios, the group returned to the holding area where they completed the post-test questionnaires.

Data Analysis

This study employed a 2 X 3 repeated measures design, drawing participants from a novice population. The independent variables were display type (HMD versus LCD flat screen) and trial (scenarios one, two and three). Additional independent variables were the participant's Immersive Tendency Questionnaire (ITQ) score and initial SSQ score. Dependent variables were performance, final SSQ scores and Presence Questionnaire (PQ) score. Display type was a between-subjects independent variable. Trial was a within-subjects independent variable.

Data was analyzed using a 2 (display type) X 3 (trial) repeated measures analysis of variance (ANOVA) for performance (number of enemy targets destroyed) by display type for hypothesis 1. For hypothesis 2, simulator sickness (total severity and sub-scale aggregate scores) was analyzed by using separate single-factor ANOVAs. A series of simple linear regression tests were conducted to determine whether a linear relationship existed between the explanatory variable (ITQ Score) and the response variables, performance and level of presence for hypothesis 3. For hypothesis 4, level of presence was analyzed using a single-factor ANOVA.

RESULTS

Hypothesis 1: Performance Effect of Different Visual Modalities

Hypothesis 1 was "*the mean door gunner performance in the HMD visual group will be equal to the mean door gunner performance in the flat screen visual group*". This hypothesis was derived from our literature review, which indicated that wearable simulation (HMDs) generally provided little to no benefit in performance when compared to traditional flat screen displays. The results of this experiment forced us to reject this hypothesis and conflicted with the results of our previous study.

ANOVA found a significant main effect of visual display on performance $F(1, 222) = 13.59$, $p = 0.00$. ANOVA also did indicate a significant main effect of scenario on performance, $F(2, 222) = 3.48$, $p = 0.03$. The interaction between visual display and trial was not significant, $F(2, 222) = 0.01$, $p = 0.99$. ANOVA was conducted at $\alpha = 0.05$.

A series of post-hoc Student's t-Tests were performed to test hypothesis 1, employing a Bonferroni correction ($\alpha = 0.01$). Post-hoc tests revealed no significant effect of visual display on performance for scenario 1 [$t(74) = 2.27$, $p = 0.03$]. There was no significant effect of visual display on performance for scenario 2 [$t(74) = 1.85$, $p = 0.07$]. There was no significant effect of visual display on performance for scenario 3 [$t(74) = 2.33$, $p = 0.02$]. However, we reject this hypothesis as a lack of support has been found.

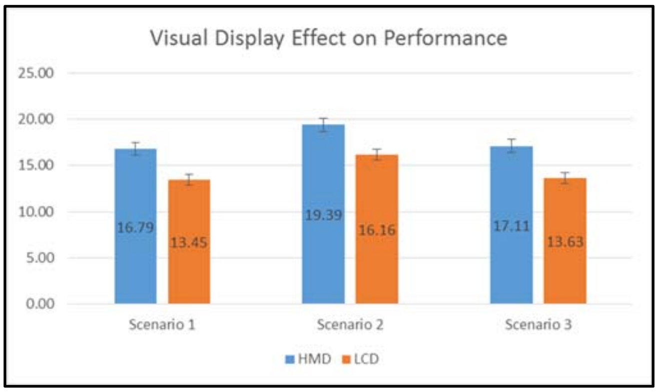

Figure 2. Visual Display Effect on Performance

There is a significant difference in the mean door gunner performance of the HMD visual group and the flat screen visual group at $\alpha = 0.05$ (Figure 2). This indicates that performance was affected by visual display.

Hypothesis 2: Level of Simulator Sickness

Hypothesis 2 was "*simulator sickness is greater in the HMD visual group than the flat screen visual group*". This hypothesis was derived from our literature review, which clearly supported that HMDs induce a greater degree of simulator sickness than traditional flat screen displays. The results of this experiment forced us to reject this hypothesis, similar to the results of our previous study.

Each of the three Simulator Sickness Questionnaire (SSQ) subscales and Total Severity scores were obtained before and after the participant's exposure to the simulator (Figure 3). A series of one-way ANOVAs ($\alpha = 0.05$), with the SSQ change scores as dependent variables and visual display as the independent variable, were conducted. No main effect of visual display was found for the Nausea subscale [F (1, 74) = 0.55, $p = 0.46$]. No main effect of visual display was discovered for the Oculomotor Discomfort subscale [F (1, 74) = 0.37, $p = 0.55$]. No main effect of visual display was discovered for the Disorientation subscale [F (1, 74) = 1.71, $p = 0.19$]. Finally, no main effect of visual display was discovered for the Total Severity score [F (1, 74) = 1.20, $p = 0.28$].

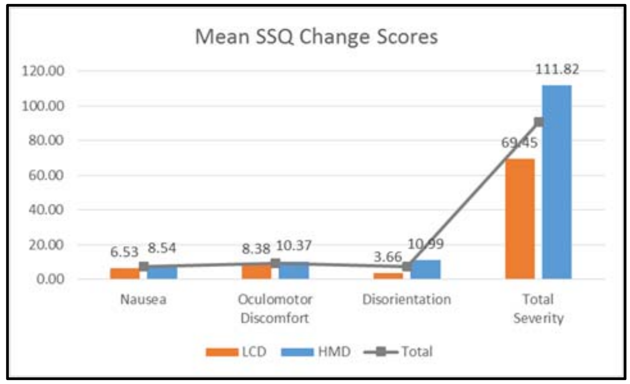

Figure 3. Mean Sickness Questionnaire Change Scores

There was no significant effect of visual display on simulator sickness. We reject this hypothesis as no support has been found. We conclude that there is no statistical difference in the SSQ subscale scores nor Total Severity scores between both visual display groups.

Hypothesis 3: Level of Immersion

Hypothesis 3 was "*there does not exist a relationship between an individual's immersive tendency score and their performance and level of presence in a Mixed Reality Aerial Door Gunnery Training environment*". This hypothesis was derived from our belief that an individual's tendency to be immersed, as measured by the ITQ, had no effect on an individual's performance nor their level of presence experienced in the simulator. The results of this experiment led us to fail to reject this hypothesis, as significant support was found. This was congruent with our prior study's findings.

The first regression test executed was all participants' performance coupled with their ITQ score (Figure 4). Performance was calculated as the average score the participant achieved over their three trials. ITQ score did not predict subject performance, $\beta = -0.02$, $t(75) = -0.47$, $p = 0.64$. ITQ score also did not explain a significant proportion of variance in performance scores $R^2 = 0.003$, $F(1, 74) = 0.22$, $p = 0.64$. Additionally, subsequent simple linear regression tests revealed that no significant linear relationship existed between an individual's ITQ score and performance when both visual treatment groups were examined separately.

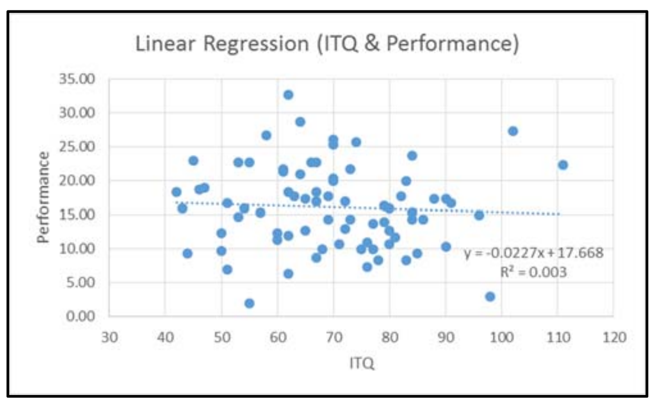

Figure 4. Linear Regression Results for ITQ and Performance Scores

The second simple linear regression test conducted was to determine if a linear relationship existed between the explanatory variable (Immersive Tendencies Questionnaire Score (ITQ)) and the response variable, Presence Questionnaire (PQ) score (Figure 5). We found no linear relationship exists. ITQ score did not predict PQ score, β = -0.07, t(75) = -0.51, $p = 0.61$. ITQ score also did not explain a significant proportion of variance in PQ scores $R^2 = 0.003$, F(1, 74) = 0.26, $p = 0.61$. Additionally, similar to performance, when segregated by visual treatment, we found no linear relationship existed between the explanatory variable (ITQ) and the response variable, Presence Questionnaire (PQ) score.

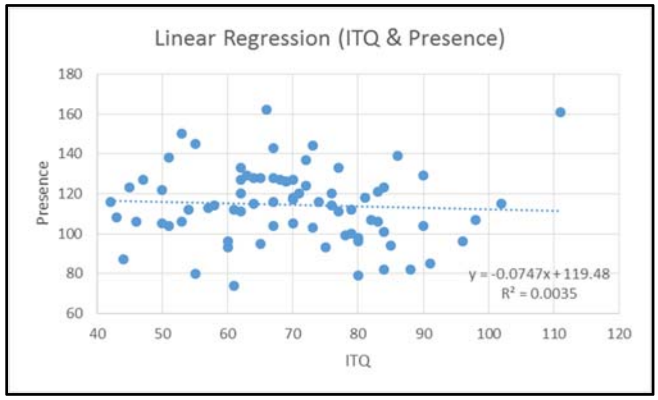

Figure 5. Linear Regression Results for ITQ and PQ Scores

We fail to reject hypothesis 3 as support has been found. The results of simple linear regression tests lead us to conclude that there was no significant linear relationship found between an individual's immersive tendency and their performance, nor their level of presence experienced.

Hypothesis 4: Level of Presence

Hypothesis 4 was "*the level of presence in the HMD visual group will be equal to the level of presence in the flat screen visual group*". This hypothesis was derived from our literature review, which, based upon recent experimentation, indicated that the degree of presence experienced in simulation can be independent of the simulator's physical immersion. In other words, higher presence has been empirically demonstrated in lower immersive settings and vice versa. The results of this experiment led us to fail to reject this hypothesis, as significant support was found. This was congruent with our prior study's findings.

Participants completed their Presence Questionnaire (PQ) after simulator exposure. The PQ provided a numerical value of the degree of presence the trainee experienced, as reported by the individual. The independent variable was visual display. No main effect of visual display was found for the Presence Questionnaire score [F (1, 74) = 3.89, $p = 0.05$] in accordance with Figure 6.

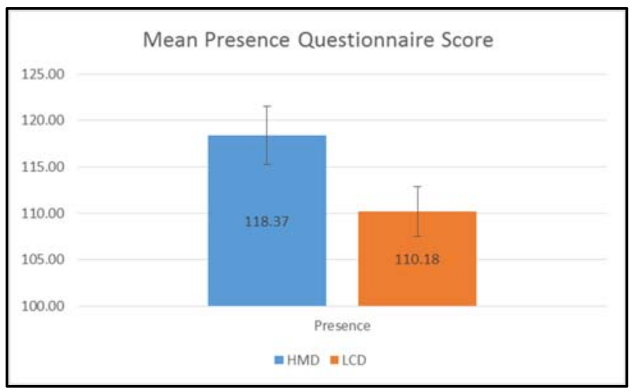

Figure 6. Mean Presence Questionnaire Scores

There was no significant effect of visual display on level of presence experienced in the simulator. We conclude that the degree of presence experienced in the simulation was statistically equal between both visual treatment groups.

DISCUSSION

Conclusions

The test results of Hypothesis 1, "*the mean door gunner performance in the HMD visual group will be equal to the mean door gunner performance in the flat screen visual group*" indicated a statistically significant difference in performance between the HMD group and the LCD group at α = .05, with HMD gunners outperforming the LCD group. The novice group's performance also did not remain constant amongst consecutive trials at α = .05. These results conflicted with our prior study whereby we determined that expert performance was not affected by visual display.

A possible explanation for the conflicting performance results amongst different expertise groups involves the spatial component of the aerial door gunnery training task. In order to be successful, participants needed to rapidly detect and acquire targets prior to engaging them with the machine gun. Target detection and acquisition, the primary spatial component of this task, was easier for participants in the HMD treatment than the LCD treatment because of both the slightly wider field of view (~8^0 in each direction) as well as the HMD's head-tracking capability. The HMD group had precious more milliseconds to detect and acquire incoming targets than the LCD group. The expert group was able to overcome that time differential, in large part due to their familiarity and experience with the training task. For the novice group, who had no prior NCM experience but did possess machine gun experience, this proved to be the primary driver behind the performance difference found between visual display groups. This finding is congruent with Knerr's (2007) determination that HMDs are slightly more effective than traditional visual displays when performing spatially oriented tasks in a virtual environment.

The test results of Hypothesis 2, "*simulator sickness is greater in the HMD visual group than the flat screen visual group*" indicated no difference in the level of simulator sickness between both visual display groups at $\alpha = .05$ and was thus rejected. We conclude that the level of simulator sickness was the same amongst both visual groups. This result mirrored our findings from the first study. Similar to the first study, the stationary nature of the task, the comparable field of views afforded by both visual treatments and the brief simulator exposure time all represent possible explanations for the equal level of SS amongst both visual treatments.

Hypothesis 3 was "*there does not exist a relationship between an individual's immersive tendency score and their performance and level of presence in a Mixed Reality Aerial Door Gunnery Training environment* ". The test results of hypothesis 3 indicated no linear relationship existed between a subject's Immersive Tendencies Questionnaire Score (ITQ) and their performance score nor their perceived level of presence. This determination is consistent with our findings from the first study, utilizing an expert population. Similar to before, we again question the utility of the ITQ as a predictive tool.

The test results of Hypothesis 4, "*the level of presence in the HMD visual group will be equal to the level of presence in the flat screen visual group*" indicated no difference in the level of presence perceived by subjects in either visual treatment at $\alpha = .05$. The lack of difference in the level of presence between both visual groups was consistent with the results from our first study. Again, we posit that the level of presence experienced by participants was driven more by the kinetic emphasis of the task being trained than the visual display being utilized.

Lessons Learned

The results of this study demonstrated that visual display had a significant effect on novice door gunners' performance in a mixed reality simulator. The novice group's result was unexpected by the authors, based upon both the literature review conducted as well as the results from our first study utilizing an expert population. This experiment would have provided more meaningful results if the participants were subsequently tested on aerial door gunnery in a live environment (using real aircraft, weaponry and ammunition) after simulator exposure. However, due to resource constraints and safety considerations, this was not possible.

Additionally, this study demonstrated that simulator sickness was statistically equal between the HMD group and the LCD flat screen group, an unexpected result. However, the Principal Investigator designed this experiment to support throughput and thus limited participating soldiers' individual exposure to the simulator to under 30 minutes. This undoubtedly affected the degree of SS experienced by participants, per the literature review. Expectedly, immersive tendency was not found to be a reliable predictor of performance nor of presence experienced by the participants. Finally, the level of presence experienced by participants did not differ by visual display group, as expected.

This study can hopefully serve as a tool in the Program Manager's toolkit in order to have a professional dialogue with his/her requirements manager and/or program proponent. Procuring dual visual solutions for one simulator not only increased the complexity of the system but also unnecessarily increased the cost. While the cost increase was not dramatic, due to the low number of manned modules in the NCM3, this may not always be the case. This study hopefully highlighted that the latest technology is not always the optimal solution. Analysis and trade-offs must be conducted and all parties must have a baseline understanding of the second and third-order effects of these decisions.

Furthermore, the results and lessons learned from this study are not unique to the military nor specific to aerial door gunnery. The most effective visual display selection is of paramount importance to the designers of driving, flying, manufacturing and countless other types of virtual simulators involving psychomotor task training. As examples, Bruzzone et al. (2011) described their development of a prototype portainer virtual simulator that will be used to train operators in port-related tasks. van Wyck & de Villiers (2009) outlined their approach to developing a virtual reality prototype for safety training in mining and drilling operations. Insinna (2014) described industry's visual solution options for a new F-35 tactical simulation. This study may have been of benefit to those authors during their respective simulation design phases, when they chose their visual solutions.

Recommendations for Future Research

In this study, we analyzed the effect different visual displays had on the performance of aerial door gunnery in a mixed reality training environment. Although it was not possible to validate the study's results in a live training environment, this would obviously represent the major recommendation moving forward. The measurement of the degree of transfer, based upon visual display used in the simulator, to the live environment, would be a worthwhile endeavor. Similar to this study, the recommendation would be for the research team to employ a between-subjects experimental design.

Other recommendations for future research would include the addition of other visual treatments, such as a wider FOV HMD and wider FOV adjustable flat screen displays, to ascertain the true impact of FOV on this particular training task. Adding an additional, discontinuous session with the participants and comparing the rate of performance decay that both visual display groups experience would present another opportunity for valuable research. Finally, the possible inclusion of a third group to this study, with a lower expertise level than the current novice bracket (perhaps from a combat support unit) would be noteworthy research as well.

ACKNOWLEDGEMENTS

The authors would again like to personally thank Ms. Belinda Lewis and the Leidos, Inc. team for their tremendous support of this research.

REFERENCES

Aegis Technologies. (2008). *Metrics for Modeling and Simulation (M&S) Investments.* Huntsville, AL: The AEgis Technologies Group, Inc.

Alexander, A., Brunye, T., Sidman, J., & Weil, S. (2005). From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games. *DARWARS Training Impact Group.* Woburn, MA.

Barnett, J. S., & Taylor, G. (2012). *How Simulator Interfaces Affect Transfer of Training: Comparing Wearable and Desktop Systems.* Ft. Belvoir, Virginia 22060: United States Army Research Institute.

Blow, C. (2012). *Flight School in the Virtual Environment.* Fort Leavenworth, Kansas: United States Army Command and General Staff College.

Borgvall, J. (2013). Generating Recommendations for Simulator Design Through Pilot Assessment of Prototype Utility. *Military Psychology*, 244-251.

Bruzzone, A., Fadda, P., Fancello, G., Massei, M., Bocca, E., Tremori, A., . . . D'Errico, G. (2011). Logistics Node Simulator as an Enabler for Supply Chain Development: Innovative Portainer Simulator as the Assessment Tool for Human Factors in Port Cranes. *Simulation*, 857-874.

Curran, M. (2003). Army Aviation Operations During Operation Iraqi Freedom. *Army Aviation*, 6-12.

Dalgarno, B., & Lee, M. (2010). What are the Learning Affordances of 3-D Virtual Environments? *British Journal of Educational Technology*, 10-32.

Harrington, M. (2011). Empirical Evidence of Priming, Transfer, Reinforcement, and Learning in the Real and Virtual Trillium Trails. *Learning Technologies IEEE Transactions on*, 175-186.

Hays, R., Jacobs, J., Carolyn, P., & Salas, E. (1992). Flight Simulator Training Effectiveness: A Meta-Analysis. *MILITARY PSYCHOLOGY*, 63-74.

Insinna, V. (2014, February). Tactical Trainer Would Teach F-35 Pilots Decision-Making Skills . *National Defense*.

Jacquet, C. (2002). *A Study of the Effects of Virtual Reality on the Retention of Training.* Orlando: University of Central Florida.

Jentsch, F., & Bowers, C. (1998). Evidence for the Validity of PC-based Simulations in Studtying Aircrew Coordination. *The International Journal of Aviation Psychology*, 243-260.

Kennedy, R., Lane, N., Berbaum, K., & Lilienthal, M. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 203-220.

Knerr, B. (2007). *Immersive Simulation Training Immersive Simulation Training.* Arlington, Virginia: United States Army Research Institute.

Marton, A. (2008). Technologies for Rotary Aviation Crew Coordination and Collective Training. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2008.* Orlando: National Training Systems Association.

McMahan, R., Bowman, D., Zielinski, D., & Brady, R. (2012). Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game. *Visualization and Computer Graphics*, 626-633.

Mikropoulos, T., & Natsis, A. (2011). Educational Virtual Environments: A Ten-Year Review of Empirical Research (1999–2009). *Computers & Education*, 769-780.

Milgram, P., & Colquhoun, H. (1999). A Taxonomy of Real and Virtual World Display Integration. In P. Milgram, & H. Colquhoun, *Mixed Reality: Merging Real and Virtual Worlds* (pp. 5-30).

Muchinsky, P. (2000). *Psychology Applied to Work.* Belmont, CA : Wadsworth.

Persky, S., Kaphinqst, K., McCall, C., Lachance, C., Beall, A., & Blascovich, J. (2009). Presence Relates to Distinct Outcomes in Two Virtual Environments Employing Different Learning Modalities. *Cyberpsychology & Behavior*, 263-268.

Santos, B. S., Dias, P., Pimentel, A., Baggerman, J. W., Ferreira, C., Silva, S., & Madeira, J. (2009). Head-Mounted Display versus Desktop for 3D Navigation in Virtual Reality: A User Study. *Multimedia Tools and Applications*, 161-181.

Scales, B. (2013, July). Virtual Immersion Training: Bloodless Battles for Small-Unit Readiness. *The Magazine of the Association of the United States Army*, pp. 24-27.

Schout, B., Hendrikx, A., Scheele, F. B., & Scherpbier, A. (2010). Validation and Implementation of Surgical Simulators: A Critical Review of Present, Past, and Future. *Surgical Endoscopy*, 536-546.

Selverian, M., & Sung, H. H. (2003). In Search of Presence: A Systematic Evaluation of Evolving VLEs. *Presence: Teleoperators & Virtual Environments*, 512-522.

Seymour, N., Gallagher, A., Roman, S., O'Brien, M., Bansal, V., Anderson, D., & Satava, R. (2002). Virtual Reality Training Improves Operating Room Performance. *Annals of Surgery*, 458-464.

Stevens, J., & Kincaid, P. (2014). Measuring Visual Modalities' Effect on Expert Performance in Mixed Reality Aerial Door Gunnery. *MODSIM World 2014.* Hampton, VA: National Training and Simulation Association.

Stewart, J., Johnson, D., & Howse, W. (2008). *Fidelity Requirements for Army Aviation Training Devices: Issues and Answers.* Arlington, Virginia: U.S. Army Research Institute.

Summers, J. (2012). Simulation-based Military Training: An Engineering Approach to Better Addressing Competing Environmental, Fiscal, and Security Concerns. *Journal of the Washington Academy of Sciences*.

Thorpe, J. (2010). Trends in Modeling, Simulation & Gaming: Personal Observatiions about the Past Thirty Years and Speculation about the Next Ten. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).* National Training Systems Association.

United States Government Accountability Office. (2013). *Army and Marine Corps Training: Better Performance and Cost Data Needed to More Fully Assess Simulation-Based Efforts.* Washington, DC: Government Accountability Office.

van Wyk, E., & de Villiers, R. (2009). Virtual Reality Training Applications for the Mining Industry. *Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa* (pp. 53-63). ACM.

Waller, D., Hunt, E., & Knapp, D. (1998). The Transfer of Spatial Knowledge in Virtual Environment Training. *Presence: Teleoperators and Virtual Environments*, 129-143.

Witmer, B., & Singer, M. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*, 225-240.