Measuring Visual Modalities' Effect on Expert Performance in Mixed Reality Aerial Door Gunnery

LTC Jonathan Stevens
Army Research Laboratory (ARL)
Orlando, FL
jonathan.stevens@us.army.mil

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

ABSTRACT

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. The goal of this first study was to evaluate visual modalities' effect on performance in mixed reality aerial door gunnery. Participants were randomly assigned to one of two visual modality treatments (flat screen or Head-Mounted Display) and executed three aerial door gunnery training scenarios in the NCM3. Independent variables were visual modality, immersive tendency and simulator sickness questionnaire scores. Dependent variables included performance, presence and simulator sickness change scores. The results of the study indicate no main effect of visual modality on performance. Both visual treatment groups experienced the same degree of presence and simulator sickness. Results of this study may challenge the commonly held notion that higher immersive simulation leads to better performance and presence.

For this initial study, participants were drawn from an expert population of qualified non-rated crew members. Subsequent studies will examine visual modalities' effect on performance in mixed reality aerial door gunnery utilizing novice participants.

ABOUT THE AUTHORS

Jonathan Stevens is a 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 21 years of military experience as an Infantry and Acquisition Corps officer. LTC Stevens is currently enrolled as a PhD student in the University of Central Florida's Modeling and Simulation graduate program.

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 Modalities' Effect on Expert Performance in Mixed Reality Aerial Door Gunnery

LTC Jonathan Stevens
Army Research Laboratory (ARL)
Orlando, FL
jonathan.stevens@us.army.mil

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

INTRODUCTION

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' recent article is an exemplar of leader bias not grounded in science (Scales, 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. One such example exists that serves as the basis for this research.

Non-Rated Crew Members in the U.S. Army & 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 parlance, meaning they cannot fly the aircraft. Some of the most critical tasks that a NCM must perform are aerial door gunnery, sling load operations and crew coordination. However, until very recently, these tasks could only be trained in a live environment (Marton, 2008). 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 critical, high risk tasks. The NCM3 is a mobile, transportable, multi-station virtual simulation device designed to support training of non-rated crew members in crew coordination, flight, aerial gunnery, hoist and sling load related tasks (Stevens & Samouce, 2011). Each single trailer NCM3 system contains two manned modules (MMs) re-configurable to either a UH-60 (Blackhawk) or CH-47 (Chinook). There are two instructor/operator stations and an integrated semi-automated forces (SAF) for modeling of threat and friendly units. An exercise record/playback capability is provided for an integrated after-action review (AAR) (Stevens & Samouce, 2011).

Program Manager's Dilemma

During the design phase of the NCM3, it was necessary for the Program Manager to utilize trade-off space. The original design did not call for the use of LCD panel screens in the trainer. However, the user community was adamant that aerial door gunnery must be capable of being trained in a non-HMD mode. The concession was that a less expensive HMD would be procured, with a narrower field of view. Thus, the negotiated trade-off resulted in a training device that was capable of supporting mixed reality aerial door gunnery training via HMD or LCD screen.

While necessary to keep the project moving forward, the above trade-off was conducted in a non-scientific manner, with no empirical knowledge of what the training effect would be. This runs contrary to Department of Defense modeling and simulation best-practice investment strategies, one of which calls for "quantifying the extent of potential investments and for identifying and understanding the full range of benefits resulting from these investments" (Aegis Technologies, 2008). However, the dual visual modality capability the trainer affords allows experimentation to be conducted that measures the effect those different visual modalities have on performance of aerial door gunnery.

BACKGROUND

Benefits of Simulation for Training

The goal of simulation for training is to provide "increased performance effectiveness at the same or lesser cost" (Orlansky, et al., 1994). Orlansky et al. (1994) describe the major advantages of using simulation for training as reduced cost, time and effort to conduct training. Militaries throughout the world will continue to expand their use of simulation for training (Lele, 2013) as will the medical community (McGaghie, Issenberg, Petrusa, & Scalese, 2010). Simulation-based training has provided many benefits to the aviation community, such as higher safety rates and improved team-based performance (Aguinis & Kraiger, 2009) (Dourado & Martin, 2013) and in some instances simulators have replaced live aircraft in training (Kincaid & Westerlund, 2009). Simulation-based training is an effective and efficient alternative to one-on-one tutoring, provided the right instructional strategy is employed (Vogel-Walcutt, Carper, Bowers, & Nicholson, 2010).

Taxonomy of Reality

Mixed reality (MR) is the space that lies between the extremes of a completely virtual environment and completely real environment in the Reality-Virtuality (R-V) continuum (Milgram & Colquhoun, 1999). Mixed reality can be further decomposed into Augmented Reality (AR) and Augmented Virtuality (AV), where augmented reality is a "type of virtual reality in which synthetic stimuli are registered with and super-imposed on real-world objects" (Sherman & Craig, 2003). In AR, the live environment is augmented with virtual data and graphics, whereby in AV, the virtual environment is augmented with real data and images.

Milgram & Colquhoun (1999) developed the global taxonomy of MR display integration (Figure 1). The three dimensions of the taxonomy consist of the R-V continuum (previously described), Congruence and Centricity. Congruence is the level of natural response shown in the user's display space and is reactive to the user's input. Higher congruence implies more direct and natural user control through the device's interface. Centricity refers to the user's viewpoint; an egocentric centricity represents a first-person view whereas an exocentric centricity represents a third-person, world-view. This study will be comparing the effect on performance between a traditional visual display (Class 1) and an immersive Head Mounted Display (HMD) (Class 2) operating in the red oval of the MR Global Taxonomy.

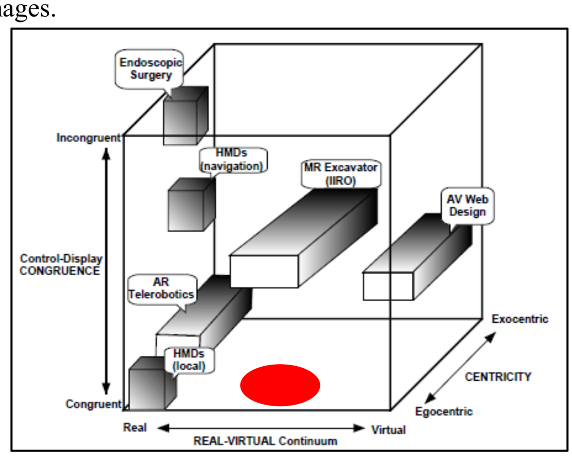

Figure 1: Global Taxonomy of MR Display Integration

Training Efficacy of Virtual Environments

The goal of a virtual simulation is generally to maximize the degree of skill transfer to the trainee. Transfer can be described as 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). Thus, the higher the degree of transfer, the more successful a training system is considered. However, measuring transfer in simulation-based training (SBT) is difficult and thus not common-place. The U.S. Army does not typically employ an objective approach to measuring transfer in simulation-based training (Insinna, 2013) and has been criticized for this practice (United States Government Accountability Office, 2013).

While difficult to measure, 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). However, it is costly, from a resource and monetary standpoint, to measure transfer of training by employing the above methodology (virtual simulator performance followed by live performance). In many cases, such as this study, it is not possible to measure the transfer of training in a live system - either due to cost, safety concerns, resource availability or numerous other constraints. Therefore, a common approach is to measure transfer, or the degree of learning, in the simulator itself.

Fidelity is defined as the degree to which the virtual environment is indistinguishable from the real environment (Waller, Hunt, & Knapp, 1998) or the degree of similarity between a simulator and the environment being simulated (Borgvall, 2013). Fidelity, in the context of simulation, can be decomposed into physical and functional fidelity. Physical fidelity is defined as "the degree to which the physical simulation looks, sounds, and feels like the operational environment in terms of the visual displays, controls, and audio as well as the physics models driving each of these variables" (Alexander, Brunye, Sidman, & Weil, 2005). Functional fidelity is defined as "the degree to which the simulation acts like the operational equipment in reacting to the tasks executed by the trainee" (Alexander, Brunye, Sidman, & Weil, 2005). Research indicates that the highest level of fidelity is not necessary, but that the simulation must possess a sufficient level of fidelity where needed to train the tasks that have been selected (Summers, 2012) (Thorpe, 2010) (Borgvall, 2013) (Jentsch & Bowers, 1998).

While the literature interchanges the concepts of immersion and presence, for this study, we define immersion as "the objective level of fidelity of the sensory stimuli produced by a virtual reality system" (Ragan, Sowndararajan, Kopper, & Bowman, 2010). Presence, on the other hand, is defined as "the subjective experience of being in one place or environment, even when one is physically situated in another" (Witmer & Singer, 1998). Thus, immersion is primarily associated with the technology of the simulation (physical) while presence is associated with the trainee's subjective experience in the simulation (mental). While it is commonly thought that the greater the degree of immersion and presence induced in the simulator is correlated with higher transfer, 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). Similar to fidelity, it is not as simple as "more is better".

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) and represents a major distraction in simulation based training. SS was first reported in 1957 in a helicopter simulation (Kaufmann, Kozeny, Schaller, Borwn, & Hitz, 2012). There is no unanimous agreement on the cause of simulator sickness (Kolasinski, 1995) (Knerr, 2007) (Johnson, 2005) (Classen, Bewernitz, & Shechtman, 2011) as multiple theories (cue-conflict, ecological theory) and beliefs exist. Generally, SS is thought to result from a conflict between the body's visual and propriocentric systems, meaning there is a disconnect in the brain's positional memory (Classen, Bewernitz, & Shechtman, 2011). Kolasinski (1995) identified potential factors involved in simulator sickness in virtual environments and posited that no single factor can be identified as the cause but rather a combination of the characteristics of the individual, simulator and task will determine the extent of simulator sickness of the operator.

Head Mounted Displays (HMDs)

HMDs are becoming more common in military virtual simulators as well as real, tactical equipment. An HMD is defined as "an image source and collimating optics in a head mount" (Melzer & Moffitt, 1997), generally composed of four major elements: an image source, relay optics, mounting platform and head-tracking capability (Bayer, Rash, & Brindle, 2009). Stereoscopic, immersive HMDs are becoming ever more prevalent in virtual simulation training. While the general goal of HMD design in simulation is to maximize both the field of view and resolution, the reality is that this is a trade-off process unique to the specific simulator being developed (Melzer, Brozoski, Letowski, Harding, & Rash, 2009) (Bowman, et al., 2012).

Prior Visual Modality Research in Virtual Training

There does not exist a tremendous body of published, empirical results derived from experimentation comparing the effect on performance of different visual modalities utilized in a virtual environment. In fact, "we are far from a complete understanding of the effects of display fidelity...because controlled experiments are difficult" (Bowman, et al., 2012). 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 modality.

METHOD

Participants

The Expert group was composed of qualified non-rated crew members (NCMs) of the United States Army, drawn from multiple aviation battalions located throughout the continental United States. The population (n = 16, M = 39.3, SD = 9.9) was 100% male and 0% female.

Experimental Objectives

Four experimental objectives were examined in this study - one primary and three sub-objectives. The primary experimental objective (hypothesis 1) was to determine if there exists an effect on performance when performing Mixed Reality Aerial Door Gunnery Training attributed to the visual modality (HMD or flat screen display) employed. The first sub-objective (hypothesis 2) was to determine if there exists a difference in the level of simulator sickness experienced when performing Mixed Reality Aerial Door Gunnery Training, based on visual modality. The second sub-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 third 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 modality.

Apparatus

Two different NCM3 training devices, at separate locations, were utilized for this experiment. The primary treatment of this experiment was visual modality. 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. Performance results were captured at the NCM3's Instructor/Operator (I/O) station.

Four questionnaires were employed for this 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 qualified NCMs. The Simulator Sickness Questionnaire (SSQ) (Kennedy, Lane, Berbaum, & Lilienthal, 1993), used in this study, 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) is an 18 question (7 point scale) self-reporting checklist that measured 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) 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 were created with the assistance of a subject matter expert (SME). A familiarization scenario was developed in order to provide participants with the opportunity to familiarize themselves with the simulator, machine gun and assigned visual modality. Three formal scenarios (first, second and third trial) were created for the conduct of the actual experiment. All scenarios emphasized kinetic engagements and were approximately five minutes in duration. Difficulty level was iteratively calibrated during design to match participants' expected gain in performance. Performance was captured at the I/O station as the number of enemy targets destroyed. Expert models were calculated, for each scenario, by SMEs' performance results.

Experimental Procedure

Prior to starting the experiment, 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. Each participant read and signed their consent forms and were subsequently randomly assigned to one of two visual modality 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.

A pilot study was conducted prior to formal data collection. Formal data collection was conducted in Orlando, FL and Eastover, SC.

Data Analysis

This study employed a 2 X 3 repeated measures design, drawing participants from the expert 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 variable. Trial was a within-subjects 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. Pearson's product–moment correlation coefficients were calculated to determine the statistical relationships among performance and level of presence and immersive tendency of participants. 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*". ANOVA found no significant main effect of visual modality on performance $F(1, 42) = 1.71$, $p = 0.20$. ANOVA also did not indicate a significant main effect of scenario on performance, $F(2, 42) = 2.91$, $p = 0.07$. The interaction between visual modality and trial was not significant, $F(2, 42) = 0.08$, $p = 0.92$. 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 ($\alpha = 0.01$). There was no significant effect of visual modality on performance for scenario 1 [$t(14) = 1.14$, $p = 0.27$]. There was no significant effect of visual modality on performance for scenario 2 [$t(14) = 0.92$, $p = 0.37$]. There was no significant effect of visual modality on performance for scenario 3 [$t(14) = 0.36$, $p = 0.72$]. We fail to reject this hypothesis as support has been found. There was no 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 not affected by visual modality.

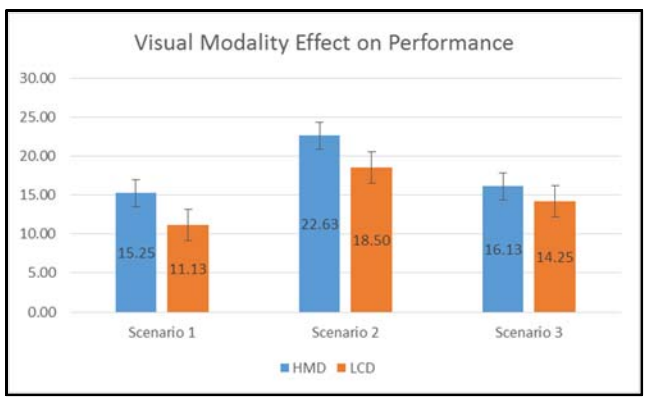

Figure 2: Visual Modality Effect on Performance

Hypothesis 2: Level of Simulator Sickness

Hypothesis 2 was "*simulator sickness is greater in the HMD visual group than the flat screen visual group*". 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. A series of one-way ANOVAs ($\alpha = 0.05$) (Table 1), with the SSQ change scores as dependent variables and visual modality as the independent variable, were conducted. No main effect of visual modality was found for the Nausea subscale [F (1, 14) = 3.32, *p* = 0.09]. No main effect of visual modality was discovered for the Oculomotor Discomfort subscale [F (1, 14) = 2.36 *p* = 0.15]. No main effect of visual modality was discovered for the Disorientation subscale [F (1, 14) = 2.33, *p* = 0.15]. Finally, no main effect of visual modality was discovered for the Total Severity score [F (1, 14) = 2.96, *p* = 0.11].

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

Table 1: ANOVA Summary for SSQ Change Scores

Source	SS	DF	MS	F	P	F Critical
Nausea	204.78	1	204.78	3.32	0.09	4.60
Oculomotor Discomfort	434.51	1	434.51	2.36	0.15	4.60
Disorientation	435.97	1	435.97	2.33	0.15	4.60
Total Severity	43919.96	1	43919.96	2.96	0.11	4.60

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*". The first regression test run was all participants' performance coupled with their ITQ score. Performance was calculated as the average score the participant achieved over their three trials. ITQ score did not predict subject performance, β = 0.15, t(15) = 0.95, *p* = 0.36. ITQ score also did not explain a significant proportion of variance in performance scores R^2 = 0.06, F(1, 14) = 0.90, *p* = 0.36. Follow-on regressions were conducted based upon visual modality and performance. Similar to the aggregated regression test, we conclude no linear relationship exists. ITQ score did not predict LCD subject performance, β = 0.11, t(7) = 0.47, *p* = 0.66. ITQ score also did not explain a significant proportion of variance in performance scores R^2 = 0.04, F(1, 6) = 0.22, *p* = 0.66. ITQ score did not predict HMD subject performance, β = 0.15, t(7) = 0.61, *p* = 0.56. ITQ score also did not explain a significant proportion of variance in performance scores R^2 = 0.06, F(1, 6) = 0.37, *p* = 0.56.

Additionally, regression was conducted to determine if a linear relationship existed between the explanatory variable (Immersive Tendencies Questionnaire Score (ITQ)) and the response variable, Presence Questionnaire (PQ) score. We found no linear relationship exists. ITQ score did not predict PQ score, β = -0.23, t(15) = -0.48, *p* = 0.64. ITQ score also did not explain a significant proportion of variance in PQ scores R^2 = 0.02, F(1, 14) = 0.23, *p* = 0.64. When segregated by visual modality, we found no linear relationship existed between the explanatory variable (ITQ) and the response variable, Presence Questionnaire (PQ) score. For the HMD visual modality, ITQ score did not predict PQ score, β = -0.31, t(7) = -0.74, *p* = 0.49. ITQ score also did not explain a significant proportion of variance in PQ scores R^2 = 0.08, F(1, 6) = 0.55, *p* = 0.49. For the LCD visual modality, ITQ score did not predict PQ score, β = -0.37, t(7) = -0.44, *p* = 0.68. ITQ score also did not explain a significant proportion of variance in PQ scores R^2 = 0.03, F(1, 6) = 0.19, *p* = 0.68.

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*". 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 modality. No main effect of visual modality was found for the Presence Questionnaire score [F (1, 14) = 1.11, *p* = 0.31] in accordance with Figure 3.

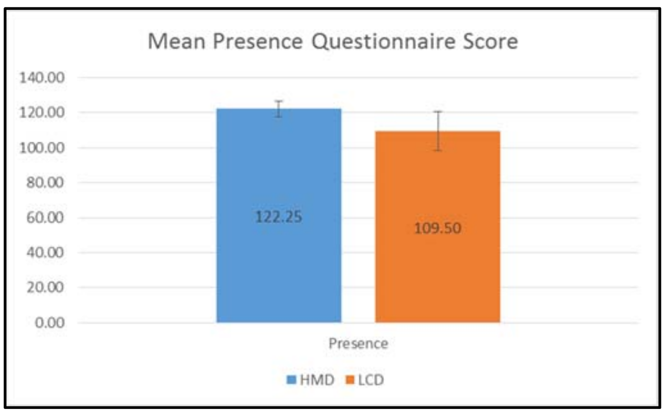

Figure 3: Mean Presence Questionnaire Scores

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 no difference in performance between the HMD group and the LCD group at $\alpha = .05$. Performance did not change from scenario to scenario, nor was a significant interaction found between visual modality and trial.

This conclusion is supported by recent research in visual modality and wearable simulation, specifically (Barnett & Taylor, 2012) and (Taylor & Barnett, 2011), where the authors found that wearable simulation was less effective than desktop-based training. While their conclusion may seem counterintuitive, the authors expand that the higher cost and higher degree of simulator sickness induced do not justify the procurement of these training systems. This study partially supports the authors' research as it was empirically found that no performance difference can be attributed to the HMD in the NCM3 for the expert population. As such, if the degree of transfer is the same between both visual modalities, then the wise choice for the Program Manager is to procure the LCD, at a cost of one-tenth the HMD's price.

This study supports prior research that concluded minimal or no benefit exists when training with a HMD versus a traditional desktop configuration. Manrique (1998), Ntuen and Yoon (2002), Knerr (2007), Singer et al. (1995) and Jacquet (2002) all reached similar conclusions with their research. Interestingly, this finding is at odds with popular opinion in the DoD simulation community.

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 modality groups. Hypothesis 2 was subsequently rejected as there was no significant difference between the HMD and LCD flat screen group SSQ Total Severity change scores at $\alpha = .05$. There also was no statistical difference in the treatment groups' Nausea, Oculomotor Discomfort or Disorientation subscale change scores. We conclude that the level of simulator sickness was the same amongst both visual groups.

Possible explanations for the equal level of SS amongst both visual treatments are the task, simulator and exposure time. The task being trained, aerial door gunnery, is stationary in nature. Door gunners are statically located at the aircraft aperture. Movement, in both visual treatments, consists of the trainee rotating his/her head to detect, acquire and engage enemy targets. There is minimal opportunity for propriocentric and visual system dissonance in this training task. Another possible explanation for the lack of SSQ difference amongst visual modality groups may be attributed to the moderate field of view of the NCM3's HMD. The NVIS nVisor MH60 has a 60 degree FOV, which is approximately equal to the 45 degree FOV that the flat screen treatment affords. This finding, however, seems to contradict recent experimentation, particularly in the area of wearable simulation (Knerr, 2007) (Barnett & Taylor, 2012). However, Arthur (2000) did find minimal differences in SS scores when he studied HMD FOVs and their effect on SS. Finally, participants' exposure to the simulator was less than 30 minutes, which has been found to decrease SS incidents.

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. Furthermore, results indicated that immersive tendency scores were not significantly correlated with performance scores nor with PQ scores.

Based on the results of this study, it is difficult to conclude the ITQ is a useful predictive tool. The ITQ was not found to be a reliable predictor of performance nor did it predict the level of presence a participant would achieve in the simulator. The authors posit that the level of immersion in the simulator was approximately equal amongst both visual treatments. The rationale for this assertion is attributed to the immersive effects of both the simulated flight helmet and the NCM3's tracking capability of the machine gun. Both of these capabilities, when coupled together, possibly mitigate any different immersive effects that the two visual modalities inherently possess.

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 recent literature. Furthermore, the authors posit that the level of presence experienced by participants was driven more by the task being conducted than the visual modality being employed. In this experiment's case, the kinetic emphasis of the three scenarios overcame any effect that the visual modality had on participants" level of presence.

Recommendations for Future Research

The results of this study demonstrated that visual modality had a minimal effect on door gunners' performance in a mixed reality simulator. This experiment would have provided more meaningful results if the participants were subsequently evaluated on aerial door gunnery in a live environment (utilizing real aircraft, weaponry and ammunition). Future research may also be conducted utilizing wider FOV HMDs in order to ascertain and compare the effect they have on performance in a mixed reality setting. Finally, as visual technology continues to advance, visual components such as Google Glass, adaptable displays and other futuristic technologies should be evaluated and researched in a similar fashion.

ACKNOWLEDGEMENTS

The authors would 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.

Aguinis, H., & Kraiger, K. (2009). Benefits of Training and Development for Individuals and Teams, Organizations, and Society. *The Annual Review of Psychology*, 451-474.

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.

Arthur, K. (2000). Effects of Field of View on Performance with Head-Mounted Displays. *Dissertation: University of North Carolina at Chapel Hill*.

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.

Bayer, M., Rash, C., & Brindle, J. (2009). Introduction to Helmet-Mounted Displays. *Helmet-Mounted Displays: Sensation, Perception, and Cognitive Issues*, 47-108.

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.

Bowman, D., Stinson, C., Ragan, E., Scerbo, S., Höllerer, T., Lee, C., . . . Kopper, R. (2012). Evaluating Effectiveness in Virtual Environments with MR Simulation. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).* National Training Systems Association.

Classen, S., Bewernitz, M., & Shechtman, O. (2011). Driving Simulator Sickness: An Evidence-Based Review of the Literature. *The American Journal of Occupational Therapy*, 179-188.

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.

Dourado, A., & Martin, C. (2013). New Concept of Dynamic Flight Simulator. *Aerospace Science and Technology*.

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. (2013, December). Army, Marine Corps Look for Better Data on Simulator Effectiveness. *National Defense*, pp. 26-28.
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.
Johnson, D. (2005). *Introduction to and Review of Simulator Sickness Research.* Arlington, VA: U.S. Army Research Institute.
Kaufmann, B., Kozeny, P., Schaller, S., Borwn, J., & Hitz, M. (2012). May Cause Dizziness: Applying the Simulator Sickness Questionnaire to Handheld Projector Interaction. *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers* (pp. 257-261). British Computer Society.
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.
Kincaid, J., & Westerlund, K. (2009). Simulation in Education and Training. *Proceedings of the 2009 Winter Simulation Conference* (pp. 273-280). Winter Simulation Conference.
Knerr, B. (2007). *Immersive Simulation Training Immersive Simulation Training.* Arlington, Virginia: United States Army Research Institute.
Kolasinski, E. (1995). *Simulator Sickness in Virtual Environments.* Alexandria: U.S. Army Research Institute.
Lele, A. (2013). Virtual Reality and its Military Utility. *Journal of Ambient Intelligence and Humanized Computing*, 17-26.
Manrique, F. (1998). *The Effects of Display Type and Spatial Ability on Performance During a Virtual Reality Simulation.* Wright Patterson Air Force Base, Ohio: Arizone State University.
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.
McGaghie, W., Issenberg, S., Petrusa, E., & Scalese, R. (2010). A Critical Review of Simulation-Based Medical Education Research: 2003–2009. *Medical Education*, 50-63.
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.
Melzer, J., & Moffitt, K. (1997). *Head-Mounted Displays: Designing for the User.* New York: McGraw-Hill.
Melzer, J., Brozoski, F., Letowski, T., Harding, T., & Rash, C. (2009). Guidelines for HMD Design. In *Helmet-Mounted Displays: Sensation, Perception and Cognition Issues* (pp. 805-848).
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.
Ntuen, C., & Yoon, S. (2002). *Assessment of Human Interaction with Virtual Environment Training Technology.* Mesa, AZ: Air Force Research Laboratory.
Orlansky, J., Dahlman, C., Hammon, C., Metzko, J., Taylor, H., & Youngblut, C. (1994). *The Value of Simulation for Training.* Alexandria: Institute for Defense Analyses.
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.
Ragan, E., Sowndararajan, A., Kopper, R., & Bowman, D. (2010). The Effects of Higher Levels of Immersion on Procedure Memorization Performance and Implications for Educational Virtual Environments. *Presence*, 527-543.
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.

Sherman, W., & Craig, A. (2003). *Understanding Virtual Reality—Interface, Application, and Design.* San Francisco, CA: Morgan Kaufmann.

Singer, M., Ehrich, J., Cinq-Mars, S., & Papin, J.-P. (1995). *Task Performance in Virtual Environments: Steroscopic versus Monoscopic Displays and Head-Coupling.* Alexandria: U.S. Army Research Institute.

Stevens, J., & Samouce, W. (2011, September). NCM3 Fielding Brief. Orlando, FL: United States Army.

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* .

Taylor, G., & Barnett, J. (2011). *Training Capabilities of Wearable and Desktop Simulator Interfaces.* Arlington, Virginia 22202: U.S. Army Research Institute.

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.

Vogel-Walcutt, J., Carper, T., Bowers, C., & Nicholson, D. (2010). Increasing Efficiency in Military Learning: Theoretical Considerations and Practical Applications. *Military Psychology*, 311-339.

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.