Impact of Visual Scene Field of View on F-16 Pilot Performance

Jamie L. Estock, Amy L. Alexander, Emily M. Stelzer, Kathryn Baughman

Aptima, Inc.

The tremendous expense and inherent dangers of training in the aircraft have led to the increased use of simulators for practicing and maintaining air combat skills; However, the advantages and disadvantages of using high or low-fidelity simulators for such training must be specified. An experiment was conducted to examine the in-simulator performance differences between pilots flying lower-fidelity simulators compared to higher-fidelity simulators. The primary difference between the two simulators is the visual scene field-of-view. Sixteen U.S. Air Force F-16 pilots flew standard training missions as an integrated team of four (a “four-ship”) with two pilots flying in the high-fidelity simulators and two pilots flying in the lower-fidelity simulators. Various subjective and objective measures were collected to assess the pilots’ ability to maintain a briefed formation. Overall, the results suggest that pilots who practice four-ship employment in the lower-fidelity simulators can perform at the same level as those who practice in the high-fidelity simulators. Future analyses should be conducted to examine the impact of simulator fidelity on other air combat skills and on training effectiveness.

INTRODUCTION

The tremendous expense and inherent dangers of “live fly” training in aircraft have led to the increasing use of flight simulators for practicing and maintaining air combat skills. High-fidelity simulators such as the Display for Advanced Research and Technology (DART) simulators in the Distributed Mission Operations (DMO) Training Research Test-bed at the Air Force Research Lab in Mesa, Arizona (AFRL/Mesa) provide very realistic visual representations of the flight environment and nearly mirror the cockpit of the F-16 aircraft. However, this approach to training is exceedingly expensive for fielding on a large scale. Such high-fidelity simulators are restricted by their size and infrastructure requirements to a fixed location, necessitating pilot travel to the simulators.

There are numerous ways to simulate aspects of flight without the expense of the high-fidelity simulators described above. For example, the civilian simulator market contains relatively inexpensive PC-based systems for training procedures and operations in Instrument Flight Rules conditions. Although there is a commonly held belief that high-fidelity simulators provide a high degree of transfer to flight, some evidence also indicates that lower-fidelity simulators can provide training benefits without the added expense and complexity of high-fidelity simulators (Wickens & Hollands, 2000). As a result, AFRL/Mesa is currently conducting research to examine the effectiveness of lower-fidelity Deployable Tactics Trainer (DTT) simulators for training air combat skills. The goal of the DTT simulators is to provide pilots with the ability to practice and maintain air combat skills in forward operating locations. Deployed pilots generally perform missions in support of real world operations and have few or no sorties/flying range space available for training opportunities. They may therefore incur deficiencies in critical mission skills needed for other theaters (Chapman, 2006). The DTT simulators were designed to address this specific issue.

The primary difference between the DART simulators and the DTT simulators is a significant reduction in the visual scene field-of-view (FOV). Specifically, the DART simulators have a 360-degree FOV visual system whereas the DTT simulator’s visual system is made up of three 30-inch Apple Cinema High Definition Liquid Crystal Displays (LCDs). Despite the benefit of increased portability offered by the DTT simulators, the majority of F-16 subject matter experts we spoke with reported that the reduced FOV could negatively impact a pilot’s ability to maintain a briefed formation, a key skill that is required for air combat and that is dependent on visual information.

This experiment was conducted to examine the utility reactions, subjective workload ratings, and in-simulator performance differences in the ability to maintain briefed formation between pilots flying lower-fidelity DTT simulators compared to high-fidelity DART simulators. We expect differences because the DTT simulator has a smaller visual scene FOV than the DART simulator. A more narrow field of view can provide a keyhole view of the world, which may limit awareness of peripheral regions of the visual scene (Wickens, Thomas, & Young, 2000; Woods, 1984). A pilot’s ability to maintain briefed formation depends on his/her ability to see his/her flight lead in his periphery. We hypothesized that:

1. Pilots flying the DTT simulators will report lower subjective ratings of the effectiveness of the simulator than pilots flying the DART simulators.
2. Pilots flying the DTT simulators will report higher levels of subjective mental workload than pilots flying the DART simulators.
3. Pilots flying the DTT simulators will show poorer performance on objective measures related to maintaining formation than pilots flying the DART simulators.

METHODS

Participants

Sixteen U.S. Air Force F-16 instructor pilots participated in the experiment. All sixteen participants were male and the majority (69%) held the rank of Captain. The participants had a mean of 1189 total F-16 hours (*SD* = 320), and an average of 94 F-16 hours in the past 6 months (*SD* = 36).

Simulators

Two different simulators were used in this research. The DART simulators are high-fidelity simulators consisting of an F-16 Block 30 aircraft cockpit with the actual F-16 aircraft controls, displays, and Operational Flight Program (OFP). The DART simulators have a 360-degree FOV visual system with 1600x1200 pixel resolution display. Figure 1 provides a view from the inside of the DART simulator.

Figure 1. Display for Advanced Research and Technology (DART) simulator.

The Deployable Tactics Trainer (DTT) simulators are lower-fidelity simulators consisting of an F-16 Block 30 aircraft shell with the actual F-16 OFP and high-fidelity aircraft stick and throttle. However, the DTT simulators only provide the essential F-16 cockpit switches on a touch screen LCD in front of the pilot. The DTT simulator's visual system is made up of three 30-inch Apple Cinema High Definition LCDs and the SDS International AAcuity® Personal Computer Image Generation System. Figure 2 provides a view of the DTT simulator.

Figure 2. Deployable Tactics Trainer (DTT) simulator.

Experimental Design

The experimental design consisted of three between-participant variables contrasting (1) utility reactions, (2) subjective workload ratings, and (3) in-simulator performance between pilots who flew the lower-fidelity DTT simulators with pilots who flew high-fidelity DART simulators. Specifically, the design focused on the impact of visual scene FOV on pilots' subjective assessments of workload and utility and in-simulator performance during air-to-air combat missions.

The design compared the utility reactions—subjective evaluation of the usefulness of the simulator as a training device—between pilots flying the DTT simulators versus the DART simulators. The utility reactions were obtained post-training via questionnaires. Specifically, the pilots were asked to rate their level of agreement with the following statements—using a one to five Likert scale from strongly disagree to strongly agree: (1) The level of visual fidelity of this simulator was effective for this DMO training, and (2) The simulator was an effective way to train me how to maintain briefed formation.

The design also contrasted the subjective workload between pilots flying the DTT simulators versus the DART simulators. The subjective workload ratings were captured via the NASA Task Load Index (TLX). This subjective workload rating procedure was developed by NASA Ames Research Center (Hart & Staveland, 1988). The NASA TLX is a multi-dimensional rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration.

Finally, the design contrasted the in-simulator performance differences between pilots flying the DTT simulators versus the DART simulators. The in-simulator performance differences were captured through the Performance Evaluation Tracking System (PETS). PETS is used by AFRL/Mesa to obtain objective, simulator-based performance data associated with the skills necessary for air-to-air combat (Schreiber, Watz, Bennett, & Portrey, 2003).

Procedures

The four-day training research experiment at AFRL/Mesa consisted of two experimental sessions per day. The participants flew standard training missions from AFRL/Mesa's DMO Training Research Syllabus as an integrated team of four (a "four-ship") with two pilots flying in the high-fidelity DART simulators and two pilots flying in the lower-fidelity DTT simulators. Prior to flying the training missions, the team members were randomly assigned to either the high-fidelity DART or low-fidelity DTT condition. During the training missions the "four-ship" was tasked with clearing all enemy aircraft from the mission space by identifying enemy aircraft, engaging those enemy aircraft, and ensuring the mission space was clear and the home base was protected. During the training missions, data were collected in real-time via PETS. After the final training mission, the pilots completed: (1) the post-training questionnaire to evaluate the usefulness of the simulator as a training device, and (2) the NASA TLX to evaluate the subjective workload rating of the training simulators in which they flew.

RESULTS AND DISCUSSION

Utility Reactions

The pilots' mean utility reactions to the visual fidelity of the simulator are presented in Figure 3. An independent samples *t* test revealed that pilots who flew the DTT simulators rated the visual fidelity significantly less effective for training (M = 2.88) than pilots who flew the DART simulators (M = 4.38, $t(14) = -3.42$, p = 0.00, d = 1.33).

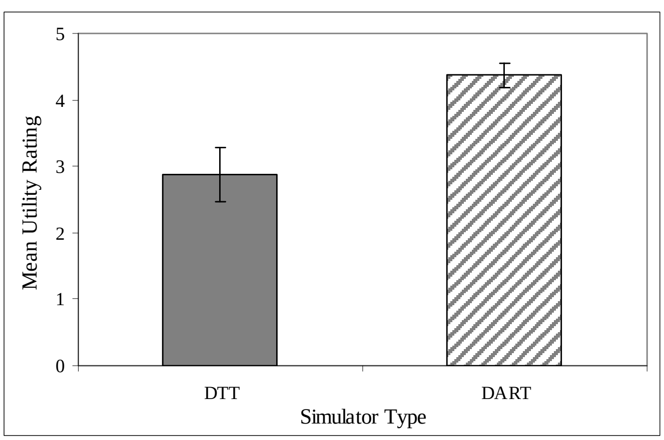

Figure 3. Mean utility reactions of visual fidelity by simulator type.

The pilots' mean utility reactions of the simulator's ability to support training on formation flying are presented in Figure 4. An independent samples *t* test revealed that pilots who flew the DTT simulators rated the simulator significantly less effective for supporting training on formation flying (M = 1.38) than pilots who flew the DART simulators (M = 3.13, $t(14) = -2.88$, p = 0.01, d = 1.29).

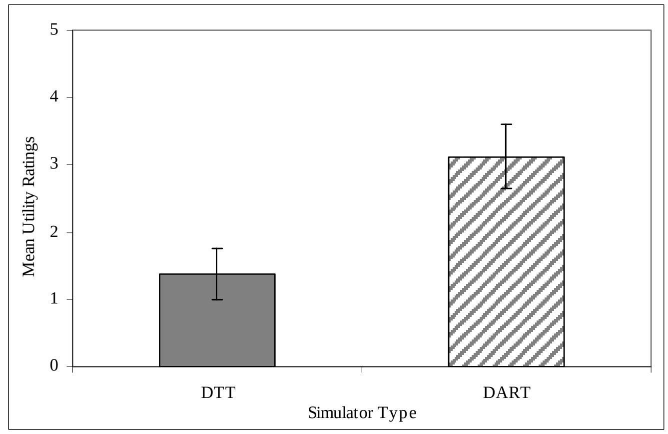

Figure 4. Mean utility reactions of simulator's ability to train maintaining formation by simulator type.

The large effect sizes—1.33 and 1.29 respectively—indicate the practical significance of the utility reaction findings. These findings provide support for hypothesis one, in which we predicted that pilots flying the DTT simulators would report lower subjective ratings of the effectiveness of the simulator compared to the pilots flying the DART simulators.

Subjective Mental Workload

Mean NASA TLX ratings are presented in Figure 5. An independent samples *t* test revealed that pilots who flew the DTT simulators reported higher mental workload (M = 75.00) than pilots who flew the DART simulators (M = 67.60, $t(14)$ = 1.73, p = 0.11, d = 0.92). While this finding is not statistically significant, the large effect size of 0.92 indicates the practical significance of this finding. This finding provides partial support for hypothesis two, in which we predicted that pilots flying the DTT simulators would report higher levels of subjective mental workload than pilots flying the DART simulators.

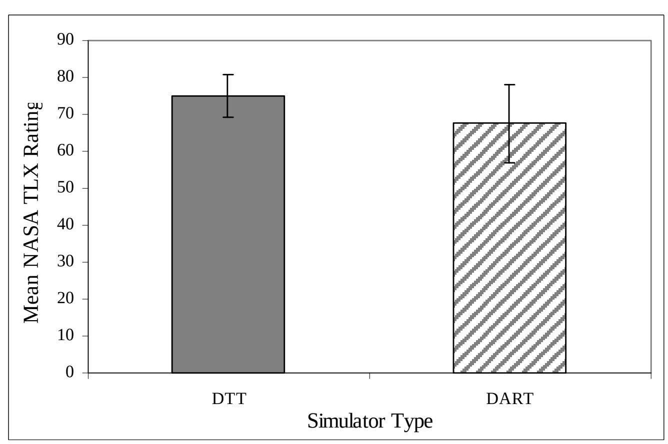

Figure 5. Mean NASA TLX ratings by simulator type.

Performance

F-16 subject matter experts identified four PETS variables that assess a pilot's ability to maintain briefed formation. Specifically, these measures include: (1) the average two-dimensional (2D) range between flight lead and wingman, (2) the average three-dimensional (3D) range between flight lead and wingman, (3) the number of mutual support violations, and (4) the total time spent in mutual support violation. A mutual support violation occurs when a pilot is outside of normal formation parameters and is less able to support his flight lead/wingman with his weapons or radar.

Because there are limitations on how PETS data may be reported (i.e., means cannot be reported in the public domain), we report percent differences between the simulator conditions, however we conducted independent samples *t* tests on the means. Percent differences in performance between pilots who flew the DTT simulators and pilots who flew the DART simulators on measures related to a pilot's ability to maintain briefed formation are presented in Table 1.

Table 1. Percent difference in performance between pilots flying DTTs versus pilots flying DARTS.

PETS variable	% Difference from DART	*t* statistic	*p* value	*d* value
Average 2D Range	-9.64%	0.16	0.88	0.10
Average 3D Range	-7.90%	0.13	0.9	0.10
Number of Mutual Support Violations	-29.04%	1.87	0.08	0.80
Total time in Mutual Support Violation	-15.59%	0.98	0.34	0.40

NOTE: The negative values presented in the % difference column reflect that the DTTs had lower values on that PETS measure.

Independent samples *t* tests revealed no difference in average 2D range, average 3D range, or the total time spent in mutual support violation within 40 nautical miles (NM) of an enemy across simulator conditions. These results fail to confirm hypothesis three, in which we predicted that pilots flying the DTT simulators would show poorer performance on objective measures related to maintaining formation compared to pilots flying the DART simulators. However, the independent samples *t* test on number of mutual support violations revealed that pilots who flew the DTT simulators had 29.04% fewer mutual support violations within 40NM of the enemy compared to pilots who flew the DARTS simulators($t(14) = 1.87$, $p = 0.08$, $d = 0.80$). While this finding is not statistically significant, the large effect size of 0.80 indicates the practical significance of this finding. These results refute hypothesis three.

CONCLUSIONS

We examined the influence of visual scene FOV differences on the utility reactions, subjective workload ratings, and in-simulator performance of F-16 pilots. The results suggest a clear discrepancy between pilots' subjective assessment of the effectiveness of simulators and the objective performance outcomes. Specifically, empirical results revealed that pilots believe that training utility is maximized when simulators are outfitted with a larger visual scene FOV. These data were replicated in open-ended responses, in which pilots reported that training was strongly inhibited by the reduced FOV of the DTT simulators. Specifically, five out of eight DTT pilots stated that FOV limitations make flying formations difficult to impossible.

However, in-simulator performance on objective, simulator-based performance measures associated with maintaining formation showed either no differences or the opposite result (i.e. number of mutual support violations). These findings may be a result of DTT pilots' heightened awareness to and compensation for the limitations of the simulators' FOV. This explanation is supported by the higher level of mental workload reported by the DTT pilots compared to the DART pilots. One DTT pilot specifically discussed the high workload related to maintaining formation in the DTT simulator. Previous research showed similar disassociation between subjective measures of mental workload and performance (Yeh & Wickens, 1988). Future analyses should be conducted to examine whether the disassociation between pilots' subjective assessment of the effectiveness of simulators and the objective performance outcomes are consistent across other air combat skills.

This finding, if shown to be consistent across other air combat skills, has important implications for verification, validation, and accreditation (VV&A) of flight simulators. Current VV&A policies and practices are not sufficient for assessing the credibility of flight simulators for training mission-level knowledge and skills (Chapman, 2006). The discrepancy between pilots' subjective assessment of the effectiveness of simulators and the objective performance outcomes demonstrates the need to collect data from multiple sources—both subjective and objective—to provide the most accurate picture of a simulator's effectiveness.

Furthermore, this research examined in-simulator performance, not training effectiveness. Future analyses should be conducted to examine the training effectiveness of simulators of varying levels of fidelity by comparing performance on pre and post-training benchmark scenarios.

ACKNOWLEDGEMENTS

We would like to thank Dr. Wink Bennett at AFRL/Mesa for funding this research. We would also like to thank Lt. Brenda Blueggel, our technical point of contact, for her support on this effort. Finally, we would like to thank Mr. Mike France, our F-16 subject matter expert, and Mr. Brian Schreiber, our PETS expert, for providing their relevant expertise toward the review of this paper.

REFERENCES

Chapman (2006). Accreditation Policy and Practice for Immersive Warfighter Simulators. In *Proceedings of the 2006 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, Florida.

Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload.* Amsterdam, The Netherlands: Elsevier.

Kraiger, K., Salas, E., & Cannon-Bowers, J. A. (1995). Measuring knowledge organization as a method for assessing learning during training. *Human Factors, 4*, 804-816.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471-499.

Schreiber, B. T., Watz, E., Bennett, W. Jr., & Portrey, A. (2003). Development of a distributed mission training automated performance tracking system. In *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation,* Scottsdale, AZ.

Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance: Third edition*. Upper Saddle River, NJ: Prentice Hall.

Wickens, C.D., Thomas, L.C., & Young, R. (2000). Frames of reference for the display of battlefield information: Judgment-display dependencies. *Human Factors*, 42(4), 660-675.

Woods, D.D. (1984). Visual momentum: A concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies*, 21, 229-244.

Yeh, Y-Y., & Wickens, C. D. (1988). The dissociation of subjective measures of mental workload and performance. *Human Factors,* 30, 111-120.