Evaluating the Impact of Dynamic Fidelity on Performance

Lisa S. Holt, Ph.D., Brian T. Schreiber, Jasmine Duran, Mark Schroeder, Ph.D.
Lumir Research Institute, Inc.
Grayslake, IL
[lisa.holt, brian.schreiber, jasmine.duran, mark.schroeder] @lumirresearch.com

ABSTRACT

Varying levels of dynamic fidelity are used to support combat mission training. At one extreme are static simulators (no motion) and at the other are costly live-fly exercises (high dynamic fidelity). Dynamic simulators employing force-cueing devices (e.g., motion platforms or motion seats) fall somewhere between. It is believed that increased dynamic fidelity in simulators will result in more effective training but empirical evidence is lacking. Numerous studies have been conducted, but the conclusions are mixed and often contradictory, with each study employing different motion cueing devices, simulation platforms, evaluation tasks, and measures of effectiveness. Both the tasks and objective performance measures used in these studies have been challenged, suggesting that they lack sensitivity to the effects of force cues.

More sensitive methodology is needed to evaluate the impact of dynamic fidelity on performance. Although there are established criteria to identify force cue-sensitive flight maneuvers, the maneuver itself does not appear to be the appropriate unit of analysis to detect performance differences. There are specific windows of time within each maneuver where the force cues play an important role in the pilot's execution and result in quantitative, measureable performance differences. Building on the research progress that has been made (defining critical phases of maneuvers and dividing them into segments), we introduce a novel unit of analysis: the force cueing sensitivity (FoCuS) window. FoCuS windows not only divide maneuvers into meaningful segments, but also include associated objective measures quantifying the role of the force cues for each segment of performance. This paper describes the application of this methodology to evaluate the impact of a motion seat on pilot performance of an advanced handling maneuver and a tactical intercept maneuver. Objective performance results contradict subjective results but illustrate the sensitivity of the method and its promise for future studies.

ABOUT THE AUTHORS

Lisa Scott Holt, Ph.D. is a Senior Scientist with Lumir Research Institute in support of the Air Force Research Laboratory, Warfighter Readiness Research Division. She is responsible for instructional design and evaluation of training technologies, and her current work investigates the relationship between simulator fidelity and training effectiveness. She received her B.S. in Physics and Mathematics from Allegheny College in 1991, and her Ph.D. in Cognitive Studies in Education from the University of Pittsburgh in 2001.

Brian T. Schreiber is CEO and Senior Scientist with Lumir Research Institute in support of the Air Force Research Laboratory, Warfighter Readiness Research Division. Mr. Schreiber received his B.S. and M.S. degrees from the University of Illinois in Champaign-Urbana. Since 1995, Mr. Schreiber has over 50 publications in books, journals, and technical reports. His work specializes in applied military research projects, particularly those involving effectiveness research and performance measurement.

Jasmine Duran is a Scientist with Lumir Research Institute in support of the Air Force Research Laboratory, Warfighter Readiness Research Division. Ms. Duran's research is focused on performance measurement and training effectiveness for military teams. She received her B.S. in Psychology and M.S. in Applied Psychology from Arizona State University in 2005 and 2010 respectively.

Mark Schroeder, Ph.D. is a Scientist with Lumir Research Institute in support of the Air Force Research Laboratory, Warfighter Readiness Research Division and also an Assistant Professor of Educational Foundations at the University of Wisconsin – Whitewater. Dr. Schroeder's research focuses on factors that create optimal learning environments and sustain motivation. He received his PhD in Educational Psychology from the University of Wisconsin – Milwaukee in 2007.

Evaluating the Impact of Dynamic Fidelity on Performance

Lisa S. Holt, Ph.D., Brian T. Schreiber, Jasmine Duran, Mark Schroeder, Ph.D.
Lumir Research Institute, Inc.
Grayslake, IL
[lisa.holt, brian.schreiber, jasmine.duran, mark.schroeder] @lumirresearch.com

INTRODUCTION

Environments of varying levels of dynamic fidelity are being used to support full combat mission training. At one extreme are static simulators (no motion) and at the other are costly live-fly exercises (high dynamic fidelity). Dynamic simulators employing various force cueing devices (e.g., motion platforms, g-suits, dynamic seats) are designed to provide environmental realism, and therefore fall somewhere between these extremes. It is believed that increased dynamic fidelity in simulators will result in more effective training but empirical evidence is lacking.

Decisions about when to incorporate force cueing in training are currently based on the valuable but subjective judgments of experienced pilots rather than on hard evidence from comparative studies. Numerous studies have been conducted, but the conclusions are mixed and often contradictory, with each study employing different motion cueing devices, simulation platforms, evaluation tasks, and measures of effectiveness. Both the tasks and objective performance measures used in these studies have been challenged, suggesting that they lack sensitivity to the effects of force cues. The desired evidence is lacking primarily because few tools or methods have been developed to track the impact of dynamic fidelity on training effectiveness.

In live flight, a pilot experiences various forces on the body that shape control behavior. The forces provide cues about the acceleration of the aircraft resulting from pilot control input, aircraft operation, and the environment (Heintzman, 1997; Szczepanski & Leland, 2000). These important cues are not present in a static simulator (where visuals provide the only cues about motion). Dynamic fidelity is achieved with force cueing devices that may use actual motion to provide the cues (e.g., centrifuges or motion platforms), or they may use non-motion-based means (e.g., g-suits and lap belts).

The Training Systems Product Group at Wright Patterson Air Force Base commissioned SIMTEC to perform a study to evaluate the need for force cueing and the effects of force cueing on training. The SIMTEC study concluded that there is little understanding of the relationship between force cueing and flight tasks, and that there is little data available for a comparative evaluation of performance in environments with different levels of dynamic fidelity (Heintzman, 1997; Jean, 2008). SIMTEC concluded the report by recommending an evaluation program to quantify the differences in performance for environments with different force cueing devices. The recommendation designated that the evaluation method must "focus upon simulator training in mission tasks where the pilot uses force cues in performance of the task in the aircraft. The method must be designed to collect both subjective (pilot opinion) and objective (measured performance) data for determining the effect of force cues and their fidelity upon pilot control strategies, performance, and training."

Such results can be used to define simulator fidelity requirements and identify the most effective compromises for updates to existing simulators. They can also be used to determine the most appropriate level of fidelity to supplement tactical in-flight pilot training as a means of off-loading flight time from aging in-service aircraft. The ultimate goal is to adeptly navigate the fidelity tradespace, optimizing training by using high fidelity environments only when they are critical.

Limitations of current approaches

Many studies were done in the 1970s to demonstrate the effects of motion platform technology on training effectiveness and essentially showed no real differences (Martin & Waag, 1978). The general conclusion was that the motion platform failed to provide any enhancements to pilot performance. Force cueing devices were then thought to provide no utility for training effectiveness. More recent literature (Heintzman, 1997; Heintzman & Basinger, 1999; Heintzman, Middendorf & Basinger, 1999; Szczepanski & Leland, 2000; Bürki-Cohen & Sparko, 2007) discount results of the motion studies of the 1970s, stating that they were based on limited motion technology, but also a limited understanding of the human perceptual system. Strachan (2001) points out that humans' perceptions of motion through the visual system happen after perceptions provided through other sensory systems. Lee (2005, p. 80) poses several possible explanations for the lack of scientific evidence including poor study methodology and the use of performance measures that may have been insensitive to the effects of the force cues. This suggests that new methodologies may actually be needed to detect performance differences.

With the advances in force cueing technology, there is now a wide variety of devices available for use in training. Among them is the ACME dynamic motion seat (DMS) which can simulate vehicle rotational motion by dynamically controlling a combination of lateral, vertical and longitudinal seat motions (Sutton, Skelton & Holt, 2010). Other available devices include centrifuges, g-suits, audio transducers and helmet loaders. With the increasing availability of force cueing devices, there is increasing use in training despite a lack of empirical evidence to guide that use.

Following their own recommendations (Heintzman, 1997), SIMTEC carried out an evaluation of the impact of force cueing devices (Heintzman et al., 1999) to help inform their efficient and effective use for training. They recognized the adaptability of pilots—if the proper cues are not available, pilots will find other cues to accomplish their tasks. This can result in use of different control strategies in the simulator. Even expert pilots are often not explicitly aware of their use of force cues. The SIMTEC study was therefore largely based on objective performance measures and also some pilot subjective perceptions.

Although the SIMTEC study did demonstrate positive effects of the dynamic seat on pilot behavior and performance, it was unable to identify the impact of other force cueing inputs. We suspect that the lack of sensitive measures is the largest contributing factor. Force cueing devices are intended to provide subtle inputs requiring focus on the specific tasks where these cues are intended to help. In order to capture performance differences, it may be necessary to focus analyses at a finer level of detail – perhaps even lower than the maneuver level.

Current performance measurement system, e.g., the Performance Evaluation Tracking System, or PETS (Schreiber, Watz & Bennett, 2003), do not collect data at the right level of detail to capture effects of force cueing. Training effectiveness studies tend to focus on mission-level outcome measures (e.g., number of kills, number of fratricides, etc.). Recent AFRL training effectiveness studies (Schreiber, Stock & Bennett, 2006) have included more detailed process measures to better understand performance. Such measures may even be inadequate to capture performance differences as the effects of force cueing can be very subtle and have an impact during very short windows of time.

Proposed measurement methodology

Based on the research we conclude that there are two essential features of a rigorous evaluation: (1) Relevant, realistic tasks that are expected to be sensitive to force-cueing, and (2) objective performance measures that are sensitive to force cueing, and specific to the windows of time in which effects are expected.

In response, we propose a new methodology which requires inspection of performance at a much finer level than reflected in most studies which tend to focus on entire maneuvers. Relevant tasks must be analyzed to identify the precise periods of time in which performance differences are expected, thus maximizing the chances of being able to detect the differences.

Figure 1 introduces the concept of a Force Cue-Sensitive (FoCuS) window. A mission refers to the full series of events in a tactical mission from takeoff to landing and a scenario represents a significant tactical portion of a mission, e.g., a 2-ship team faces two enemy strikers while defending its base. A maneuver is a specific flight action/profile executed during the execution of a mission, e.g., shot and exit. Multiple maneuvers occur within a scenario. FoCuS windows are the segments of a maneuver where differences in performance are anticipated with and without force cues, e.g., 'turning out.' Each maneuver will have multiple FoCuS windows. Finally, there are force cue-sensitive measures relevant to each FoCuS window that will reveal the objective differences in performance.

All aircraft are subject to several universal dynamics of flight that are either control-induced by the pilot or atmosphere-induced (e.g., wind, turbulence). These forces are generally referred to as forces of thrust, roll, yaw, and pitch. As such, pilots (to various extents) rely on these same forces as cues when flying to be consistent and smooth in control of the aircraft.

Flying without these subtle force cues can result in slightly larger, more variable control inputs by the pilot. The force cues allow the pilot to 'sense' the continuous and small deviations in aircraft dynamics by providing subtle force cueing input to the pilot about changes in the aircraft's state. When force cues are present, the pilot has more feedback, via more sources, making it easier to make control adjustments during flight. The result is less variable and smoother execution.

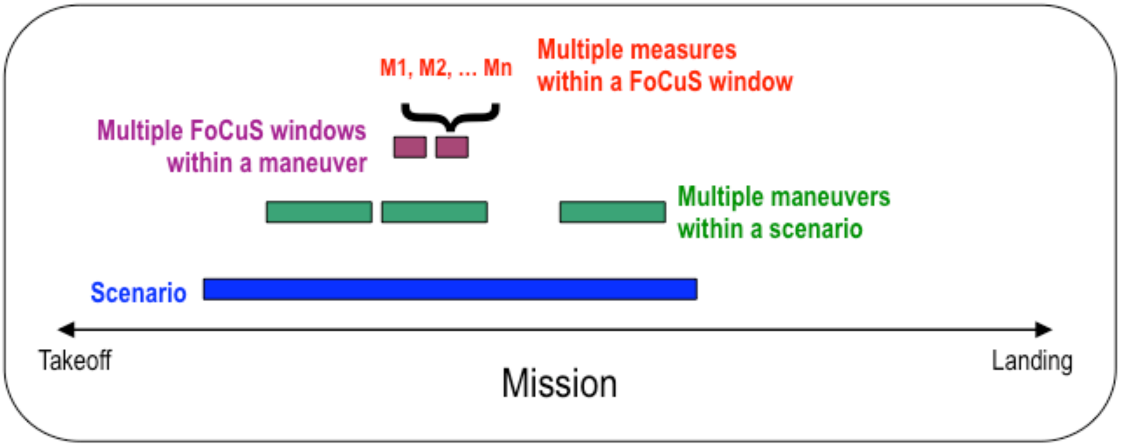

Figure 1. FoCuS window methodology

CURRENT STUDY

As part of a Cooperative Research and Development Agreement (CRADA) between ACME Worldwide Enterprises and the Air Force Research Laboratory in Mesa AZ, a study was conducted to assess the impact of the ACME DMS (ACME, n.d.) on the training capabilities of two F-16 simulators and the impact on pilot performance. A DMS was installed in a low-fidelity Deployable Tactical Trainer (DTT) and a high-fidelity Mobile Modular Display for Advanced Research and Training (M2DART).

Force cue-sensitive tasks

Industry practitioners (e.g., Heintzman et al., 1999) have specifically examined some maneuvers that may be most sensitive to force cueing, deriving a list that includes, as examples: Closing on another vehicle; maintaining separation on other vehicle; maintaining low altitude; performing ridge crossings; avoiding threats; pulling up into pop-up; tracking airborne target; tracking ground target; pulling out after ground attack; landings; air-to-air gun employment; flying formation; and pitch/roll captures.

As part of a Phase I SBIR contract, we conducted a workshop with F-16 subject matter experts (SMEs) to identify a set of potentially cue-sensitive maneuvers. Performance of these maneuvers should be dependent on the presence or absence of force cues—the force cues provide the pilot with valuable input that ultimately improves their performance. A number of tactical and administrative maneuvers were identified during the workshop.

The Horn Awareness and Recovery Training Series (HARTS) nose-high recovery and Tactical Intercept (TI) maneuvers were selected for analysis in this study for two important reasons:

- SMEs identified performance of these maneuvers as being strongly dependent on force cues.
- Relevant performance data was readily available (through PETS) for these maneuvers, while others would require programming of new measures.

The HARTS maneuver is an advanced handling characteristics exercise used to train pilots to recover from high-pitch attitude and slow airspeed conditions. To execute this maneuver, the pilot begins with wings level, climbs at 60˚ pitch until airspeed reaches a specified minimum. This puts the pilot in position to practice recovery. To recover, the pilot unloads the aircraft and rolls to the nearest horizon. As a result of the recovery maneuver, the pilot is able to regain airspeed and return to a wings-level position.

Figure 2 shows the HARTS maneuver and identifies the important FoCuS windows in which different effects of force cues are expected. The reorientation FoCuS window is the most critical as this is where the realistic piloting actually occurs. The other windows are essentially setup for and transition out of the advanced handling portion of the maneuver.

The TI maneuver is a basic tactical combat maneuver. As shown in Figure 3, the pilot approaches enemy aircraft, fires a radar-guided missile at the enemy, offsets the aircraft to maximize separation distance between his aircraft and the enemy, and then leaves once the missile no longer requires support.

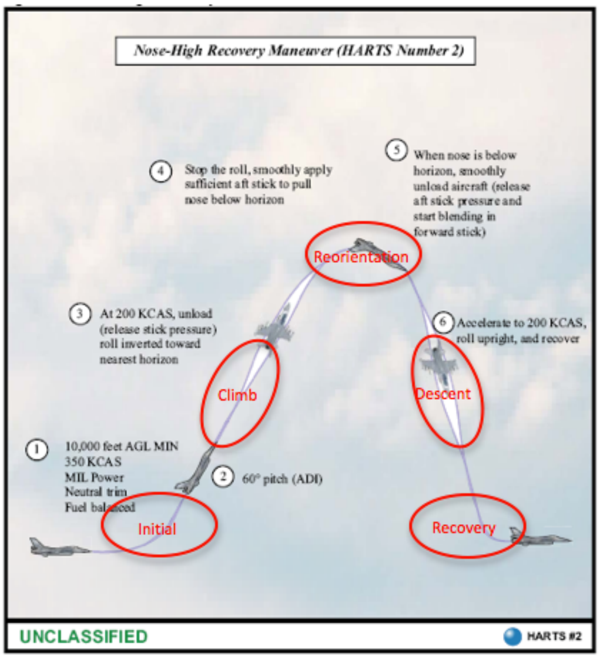

Figure 2. HARTS maneuver FoCuS windows

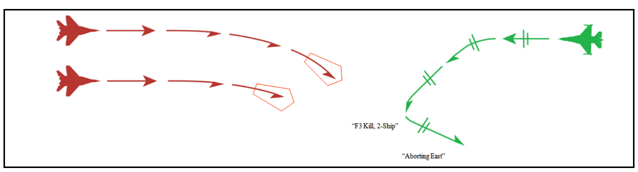

Figure 3. TI maneuver

There are two critical FoCuS windows for the TI maneuver. The crank FoCuS window (from missile launch to offset) and the out FoCuS window (during execution of the exiting turn). Effective performance during these periods is essential to the successful completion of the maneuver, resulting in a larger separation between the pilot and any surviving enemy aircraft. PETS does not currently collect the data to identify the crank window, therefore our analyses concentrated on the out window.

Study design

In Fall 2009, 12 F-16 pilots flew a variety maneuvers in the DTTs. Study tasks were balanced and randomized using a within-subjects design so that each pilot flew all tasks with and without the DMS. In Spring 2010 the study was replicated in the M2DART using 12 different pilots. As shown in Table 1, the resulting study design includes platform (DTT vs. M2DART) as a between-subjects factor and motion seat (with vs. without DMS) as a within-subjects factor.

Table 1. Study design

	With DMS	Without DMS
DTT	12 pilots x 3 maneuvers = 36	12 pilots x 3 maneuvers = 36
M2DART	12 pilots x 3 maneuvers = 36	12 pilots x 3 maneuvers = 36

Each pilot executed each maneuver three times in each condition resulting in a total of 144 HARTS maneuvers and 144 TI maneuvers for analysis.

Participants

Study participants were all male operational F-16 pilots from Luke and Davis-Monthan AFB. All 12 of the pilots in the DTT study were instructor pilots, as were 10 of the 12 pilots in the M2DART study.

The pilot groups were roughly equivalent in experience with the exception that pilots flying the M2DART were significantly older, $t(12.65) = 2.44$, $p = 0.03$, and had significantly more live flight training in the past 6 months, $t(21) = 2.21$, $p = 0.04$. (See Table 2.)

Table 2. Pilot demographics

	DTT	M2DART
Mean age	31.42 (2.19)	37.25 (7.99)
Mean yrs service	10.21 (3.42)	14.29 (8.21)
Mean F-16 hours	930.00 (226.23)	1182.08 (1613.34)
Mean combat hours	155.42 (167.57)	129.58 (136.36)
Live flight hours in past 6 months	16.61 (4.79)	55.10 (7.93)
Standard deviations in parentheses.		

Data collected

Surveys were administered to compare the training capabilities of the simulators with and without the DMS, and to asses the perceived impact of and general reactions to the DMS. Objective performance data was collected (using PETS) to determine the quantitative impact of the DMS on pilot performance.

Subjective simulator capability was assessed using a fidelity survey instrument (Schreiber, Bennett & Gehr, 2006). For each F-16 mission essential competency (MEC) experience (Colegrove & Alliger, 2002) and emergency procedure (EP), participants rated the extent to which they were able to gain the experience in the simulator. The rating scale ranged from 1 (capability exists but is very poor) to 5 (capability exists and is very good). A rating of 0 (capability does not exist) was also an option. Participants provided two ratings for each experience: one for the simulator with the DMS, and one without the DMS.

As a more sensitive measure, participants also rated the impact of the motion seat on each experience. The rating scale ranged from -2 (DMS is very detrimental) through +2 (DMS is very beneficial).

Objective performance data was collected by running PETS on the logfiles from the studies (each maneuver was recorded during the study). The performance measures collected were selected based on SME predictions of anticipated performance differences.

For the HARTS maneuver, general measures of roll, pitch, and g-load were collected at a rate of 6 Hz (6 times per second), creating an extensive log of measures across the HARTS maneuver. In our force cue workshop, SMEs identified roll, pitch and g-load as the critical elements of the flight profile for this maneuver.

Additional general measures were computed using the raw measures. Measures of smoothness and variance were calculated for each of the raw measures (e.g., smoothness of roll, variance of roll, smoothness of g-load, variance of g-load, etc.).

Smoothness was computed as dx/dt, rate of change of G-load, rate of change of pitch, etc. This measure of smoothness assesses the degree to which a pilot over controls during a maneuver, essentially reflecting when pilots use more rapid, abrupt control inputs or are 'less smooth' in their execution, thus lower smoothness scores represent better performance.

Equation 1 shows how variance was computed for each measure:

$$Var(Y) = \sum \frac{\left(Y - \overline{Y}\right)^2}{N - 1} \quad (1)$$

Y represents each instance of a measure (e.g., airspeed) for a particular maneuver, $\overline{Y}$ represents the mean value of the measure for the maneuver, and N is the number of data points for the maneuver. The variance measure represents the dispersion of scores for each given measure, providing information about variation in pilot control, with lower variance representing better pilot control.

Measures of smoothness and variance for each measure were averaged across the three HARTS maneuvers flown by each pilot in each platform/DMS condition.

Maneuver-specific measures were also calculated for the HARTS maneuver (again based on SME input from the workshop). Maximum g-load, and minimum airspeed were calculated for each HARTS maneuver.

For the TI maneuver, general measures of roll, pitch, airspeed, altitude, and g-load were collected. Smoothness and variance were calculated for each of these measures (in the same manner as for the HARTS maneuver). The maneuver-specific measures calculated for each TI maneuver included change in airspeed, change in altitude, and proportion of time spent above and below the optimal g-load range.

Hypotheses

In addition to higher subjective capability ratings with the DMS, we expected the following objective performance differences with the DMS:

- Smoother, less variable pilot control of the aircraft for both maneuvers (as measured by smoothness and variance calculations for roll, pitch, g-load, airspeed and altitude).
- Higher minimum airspeed for the HARTS maneuver (more effective execution will result in less loss of airspeed).
- Lower maximum g-load for the HARTS maneuver (excessive g-load results in less effective execution of the reorientation)
- Smaller changes in airspeed and altitude for the TI maneuver (more effective execution of the maneuver results in less loss of energy, as indicated by airspeed and altitude).
- Less time spent above and below optimal g-load range for the TI maneuver. (Excessive g-load or too little g-load results in a less effective maneuver, i.e., less separation between pilot and enemy).

We also expected that the effects of the DMS may be more pronounced in the DTT platform than the M2DART. The M2DART has 360˚ visuals, thus providing the pilot with more cues than the DTT.

SUBJECTIVE CAPABILITY RESULTS

Combining training capability rating results from the DTT and M2DART studies, the DMS was shown to significantly improve the training capability for 12 of the 70 air-to-ground (A/G) experiences, 4 of the 55 suppression of enemy air defense (SEAD) experiences, two of the 44 air-to-air (A/A) experiences, and one of the 27 emergency procedures (EP).

The impact ratings revealed a consistent beneficial impact of the DMS across the experiences. The DMS was not rated as having a detrimental impact for any of the MEC experiences or EPs. The ratings showed a statistically significant beneficial impact for 29 of the 70 A/G experiences, 12 of the 55 SEAD experiences, 21 of the 44 A/A experiences, and 8 of the 27 EPs.

In responses to open-ended questions on the survey, pilots reported improved energy management, pitch awareness, roll awareness, airspeed awareness and decreased need to monitor the G-meter as the most beneficial training aspects of the DMS. These results are consistent with pilot ratings of the impact of the dynamic motion seat on various flight phases. Pitch and roll awareness, G-onset, continuous G-monitoring, and energy management all received average ratings greater than four on a scale of five and were statistically significant.

Although pilot perceptions of simulator training capabilities and of the impact of the DMS are important, it is more important to show the resulting impact on performance.

PERFORMANCE RESULTS - HARTS

Although performance differences were expected within the reorientation FoCuS window of the HARTS maneuver (Figure 2), we were unable to isolate this window during data analysis. Extensive computational support would be required to identify the precise point where the aircraft began its roll from a position of 60˚ pitch. Despite this drawback, we opted to carry out the analysis using the data from across the entire maneuver as a point of comparison with our more sensitive FoCuS window methodology (which was successfully implemented for the TI maneuver).

Due to technical difficulties, data were not collected for two DTT pilots (one flying with the DMS, and the other flying without the DMS). In addition one smoothness of pitch measure was identified as an outlier (greater than 3 standard deviations from the mean) and removed for one of the M2DART pilots flying without the DMS. This pilot's average smoothness of pitch score was therefore based on the data from two HARTS maneuvers rather than three.

To determine the effects of force cues on the general and maneuver-specific measures, a 2X2 (platform X seat) mixed ANOVA was conducted for each. Platform is the between-subjects variable and motion seat is the within-subjects variable. Means and standard deviations associated with main effects are summarized in Table 3. Main effect and interaction results are summarized in Table 4.

General measures

There was no significant main effect of seat for smoothness of roll, variance of roll, smoothness of pitch, variance of pitch, or variance of g-load. Pilots flying with the DMS did not perform significantly different than pilots flying without the DMS in these cases. There was however a significant main effect of seat for smoothness of g-load. As shown in Figure 4, pilots flying with the DMS had significantly less erratic changes in g-load (lower smoothness scores) than pilots flying without the DMS.

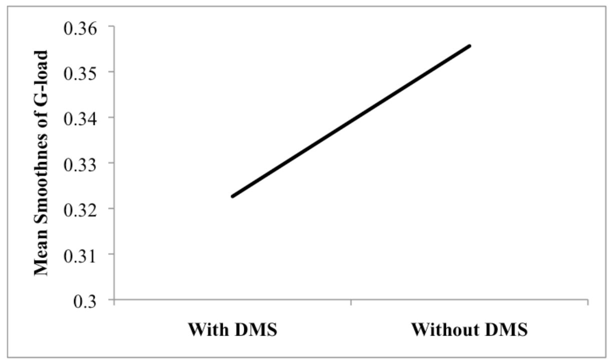

Figure 4. Main effect of motion seat for smoothness of g-load

There was no significant main effect of platform for any of the general measures (smoothness and variance of roll, pitch and g-load). M2DART pilots did not perform significantly different than DTT pilots.

There was no significant seat-platform interaction for smoothness of roll, variance of roll, smoothness of pitch, smoothness of g-load or variance of g-load. There was however a significant seat-platform interaction for variance of pitch. As shown in Figure 5, M2DART pilots had less variance in pitch when flying *with the DMS* whereas DTT pilots had less when flying *without the DMS*.

Follow-up contrasts revealed that performance difference with and without the DMS was significant for the M2DART, $t(20) = 2.56$, $p = 0.02$, $\eta^2 = 0.25$. The performance difference was not significant for the DTT. Cell means are summarized in Table 5.

Table 3. Means and standard deviations for HARTS maneuver measures

	Seat		Platform	
	With DMS	**Without DMS**	**M2DART**	**DTT**
General measures				
Smoothness of roll	40.61 (8.82)	43.15 (15.00)	43.99 (10.01)	39.34 (14.31)
Variance of roll	8387.19 (1322.33)	8711.14 (1376.04)	8695.80 (1152.49)	8373.21 (1555.17)
Smoothness of pitch	5.78 (0.91)	6.01 (0.94)	5.86 (1.06)	5.94 (0.76)
Variance of pitch	1571.88 (313.78)	1625.98 (326. 28)	1622.24 (317.10)	1570.96 (323.96)
Smoothness of g-load	0.32 (0.05)	0.36 (0.07)	0.35 (0.07)	0.33 (0.05)
Variance of g-load	1.27 (0.25)	1.33 (0.31)	1.34 (0.34)	1.25 (0.19)
Specific measures				
Minimum airspeed	202.16 (9.40)	203.25 (10.15)	205.19 (9.70)	199.71 (9.01)
Maximum g-load	3.66 (0.41)	3.76 (0.47)	3.75 (0.52)	3.66 (0.32)

Table 4. Main effects and interactions for HARTS maneuver measures

	Main effect of seat			Main effect of platform			Seat-platform interaction		
	$F(1,20)$	p	η^2	$F(1,20)$	p	η^2	$F(1,20)$	p	η^2
General measures									
Smoothness of roll	0.57	0.46	0.03	1.19	0.29	0.06	0.47	0.50	0.02
Variance of roll	0.74	0.40	0.04	0.48	0.50	0.02	0.24	0.63	0.01
Smoothness of pitch	2.77	0.11	0.12	0.04	0.85	< 0.01	0.42	0.53	0.02
Variance of pitch	1.05	0.32	0.04	594.41	0.70	0.22	5.89	**0.03**	0.22
Smoothness of g-load	4.79	**0.04**	0.19	0.58	0.45	0.03	0.37	0.55	0.02
Variance of g-load	0.86	0.37	0.04	0.70	0.41	0.03	0.86	0.51	0.02
Specific measures									
Minimum airspeed	0.18	0.64	0.01	2.27	0.15	0.10	5.58	**0.03**	0.22
Maximum g-load	1.53	0.23	0.07	0.30	0.59	0.02	0.36	0.56	0.02

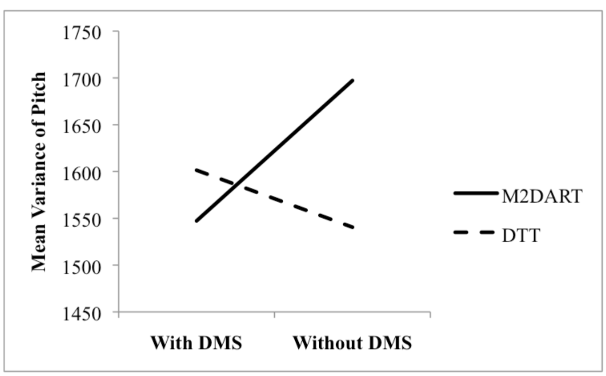

Figure 5. Interaction for variance of pitch

Maneuver-specific measures

There was no significant main effect of seat or platform for minimum airspeed or maximum g-load. Pilots flying with the DMS did not perform significantly different than pilots flying without the DMS. Similarly, M2DART pilots did not perform significantly different than DTT pilots.

Table 5. Cell means and standard deviations for HARTS variance of pitch

	M2DART	**DTT**
With DMS	1547.28 (311.87)	1601.39 (330.28)
Without DMS	1697.20 (317.38)	1540.52 (332.29)

There was no significant seat-platform interaction for maximum g-load, but the interaction was significant for minimum airspeed. As shown in Figure 6, M2DART pilots had a higher minimum airspeed (better performance) *without the DMS,* whereas DTT pilots had a higher minimum airspeed *with the DMS*.

Follow-up contrasts revealed that performance difference with and without the DMS was significant for the M2DART, $t(20) = 2.42$, $p = 0.03$, $\eta^2 = 0.23$. The performance difference was not significant for the DTT. Cell means are summarized in Table 6.

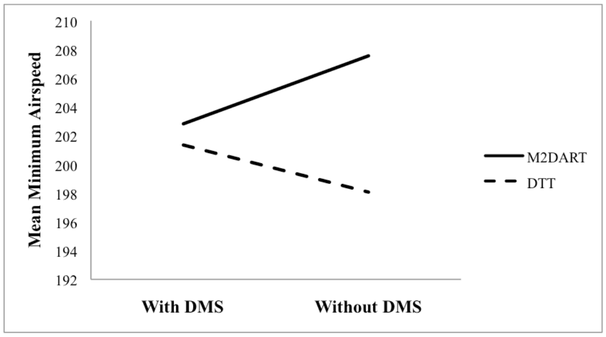

Figure 6. Interaction for minimum airspeed

Table 6. Cell means and standard deviations for HARTS minimum airspeed

	M2DART	DTT
With DMS	202.83 (10.00)	201.35 (9.09)
Without DMS	207.56 (9.19)	198.07 (9.09)

PERFORMANCE RESULTS - TI

As mentioned previously, the crank and out FoCuS windows are the most critical segments of the TI maneuver. Identification of the crank window was not possible due to lack of relevant data in PETS. We were however able to identify the start and end of the out portion of the maneuver and collect PETS measures during that time (vs. the entire maneuver as was done in the case of the HARTS maneuver). The TI analyses therefore focus exclusively on the out portion of the maneuver. Successful implementation of the FoCuS window methodology will result in more sensitive measures, thus the TI maneuver results should carry more weight than those from the HARTS maneuver. There was no missing data for the TI maneuver.

To determine the effects of force cues on the general measures, a 2X2 (platform X seat) mixed ANCOVA was conducted for each measure. Platform is a between-subjects variable and motion seat is a within-subjects variable. Pilot experience was added as a covariate in these analyses to help increase analytic sensitivity. In all cases, slopes between platforms were homogeneous and experience was neither a significant predictor of performance as a main effect nor did it interact with any other variable.

General measures

There was no significant main effect of seat or platform for any of the general measures. Pilots flying with the DMS did not perform significantly different than pilots flying without the DMS. Similarly, M2DART pilots did not perform significantly different than DTT pilots.

There was no significant seat-platform interaction for any of the general measures except variance of g-load. As shown in Figure 7, DTT pilots had less variance in g-load (better performance) *with the DMS*, whereas DART pilots performed better *without the DMS.*

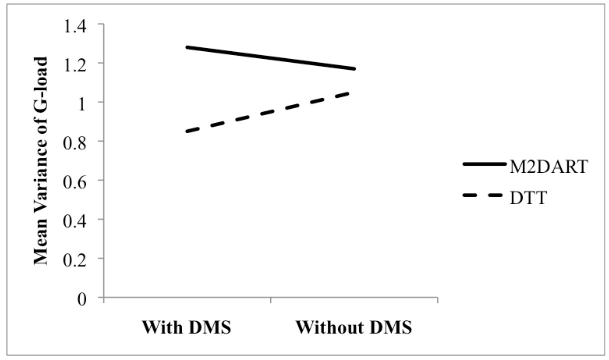

Figure 7. Interaction for variance of g-load

Follow-up contrasts revealed that the performance differences were not significant for the M2DART or the DTT. Cell means are summarized in Table 7

Table 7. Cell means and standard deviations for TI variance of g-load

	M2DART	DTT
With DMS	1.28 (0.81)	0.85 (0.45)
Without DMS	1.17 (0.79)	1.05 (0.62)

Maneuver-specific measures

There were no significant main effects of seat or platform for any maneuver-specific measures. There were also no significant seat-platform interactions.

CONCLUSIONS

An important goal of this study was not only to investigate the effects of dynamic fidelity on performance, but also to illustrate the utility of the FoCuS window methodology.

The HARTS maneuver results revealed a significant main effect of the DMS for only one variable (smoothness of g-load) in which performance was significantly better with the DMS than without the DMS across both platforms. Results also revealed two significant seat-platform interactions, one in which DART pilots performed significantly better *with the DMS* (variance of pitch), and one in which DART pilots performed significantly better *without the DMS* (minimum airspeed).

Table 8. Means and standard deviations for TI maneuver measures

	Seat		Platform	
	With DMS	**Without DMS**	**M2DART**	**DTT**
General measures				
Smoothness of roll	12.75 (3.00)	13.40 (2.90)	13.04 (3.06)	13.11 (2.89)
Variance of roll	1124.75 (677.85)	1009.44 (451.65)	1095.54 (646.16)	1038.64 (501.20)
Smoothness of pitch	4.46 (1.57)	4.39 (1.05)	4.32 (1.31)	4.53 (1.35)
Variance of pitch	155.49 (116.51)	157.42 (93.94)	140.78 (101.90)	172.13 (107.26)
Smoothness of g-load	0.61 (0.18)	0.65 (0.21)	0.66 (0.22)	0.59 (0.16)
Variance of g-load	1.07 (0.67)	1.11 (0.54)	1.23 (0.78)	0.95 (0.54)
Smoothness of airspeed	12.52 (3.90)	11.49 (1.86)	12.05 (2.18)	11.96 (3.80)
Variance of airspeed	1705.00 (1634.01)	2029.23 (1179.45)	2212.49 (1651.35)	1521.74 (1067.72)
Smoothness of altitude	415.78 (104.33)	418.85 (73.99)	410.97 (94.76)	423.67 (85.45)
Variance of altitude (divided by 1,000,000)	6.05 (2.48)	6.23 (3.50)	5.99 (3.68)	6.29 (2.20)
Specific measures				
Proportion of time below optimal g-load	0.33 (0.22)	0.33 (0.18)	0.31 (0.18)	0.35 (0.22)
Proportion of time above optimal g-load	0.09 (0.13)	0.12 (0.14)	0.13 (0.14)	0.09 (0.13)

Table 9. Main effects and interactions for TI maneuver measures

	Main effect of seat			Main effect of platform			Seat-platform interaction		
	$F(1,21)$	p	η^2	$F(1,21)$	p	η^2	$F(1,21)$	p	η^2
General measures									
Smoothness of roll	0.18	0.68	<0.01	0.17	0.68	<0.01	3.30	0.08	0.14
Variance of roll	0.15	0.70	<0.01	0.76	0.39	0.03	0.35	0.56	0.02
Smoothness of pitch	0.07	0.80	<0.01	0.02	0.89	<0.01	0.44	0.52	0.02
Variance of pitch	0.55	0.47	0.02	0.24	0.63	0.01	0.11	0.75	<0.01
Smoothness of g-load	<0.01	0.96	<0.01	1.42	0.25	0.06	0.83	0.37	0.04
Variance of g-load	1.35	0.26	0.04	1.26	0.28	0.05	5.82	**0.03**	0.19
Smoothness of airspeed	1.31	0.27	0.06	0.21	0.65	0.01	0.02	0.89	<0.01
Variance of airspeed	0.08	0.78	<0.01	1.60	0.22	0.07	4.00	0.06	0.16
Smoothness of altitude	0.15	0.70	<0.01	0.07	0.80	<0.01	0.37	0.57	0.02
Variance of altitude	1.90	0.18	0.07	0.09	0.77	<0.01	1.58	0.22	0.06
Specific measures									
Change in airspeed	0.21	0.65	0.01	<0.01	0.95	<0.01	0.90	0.36	0.04
Change in altitude	1.53	0.23	0.06	0.02	0.89	<0.01	1.96	0.18	0.08
Proportion of time below optimal g-load	0.37	0.55	0.01	0.10	0.76	<0.01	3.69	0.07	0.15
Proportion of time above optimal g-load	0.15	0.70	<0.01	0.71	0.41	0.03	0.44	0.52	0.02

Recall that the HARTS analysis was done at the maneuver level, i.e., measures were collected across the entire maneuver. This is likely to have washed out any effects of the DMS, as performance differences were only expected during the reorientation portion of the maneuver. As a result, the inconsistent HARTS results should not carry much weight.

The FoCuS methodology was implemented in the analysis of the TI maneuver in an attempt to strengthen the sensitivity of the analysis. Despite the increased sensitivity, the TI maneuver results revealed no significant main effects of the DMS. Additionally, the one significant seat-platform interaction showed no significant performance differences for either platform.

Although pilots reported improved energy management, pitch awareness, roll awareness, airspeed awareness and decreased need to monitor the G-meter as the most beneficial training aspects of the DMS, there were no objective differences in their performance.

The adaptability of pilots may have contributed to the lack of results in this study. Pilots may (unknowingly) adopt new control strategies to compensate for the lack of realistic cues, and as a result the addition of the DMS may have no impact on their performance.

Additional maneuvers need to be analyzed using the FoCuS window methodology before broad conclusions can be drawn about the impact of dynamic fidelity on performance, but the current study revealed interesting initial results. Future studies should not only examine a larger set of maneuvers, but should also consider using less experienced pilots. The force cues may have a larger impact on pilots whose control strategies are not as well established.

ACKNOWLEDGEMENTS

This work was partially supported by USAF Phase I SBIR Contract number FA8650-09-M-6979 and the Cooperative Research and Development Agreement (CRADA) between ACME Worldwide Enterprises and the Air Force Research Laboratory in Mesa AZ. The authors wish to extend special thanks to Dr. Winston Bennett Jr. for his support of this important work.

REFERENCES

Bürki-Cohen, J. and Sparko, A., (2007) *Training Value of a Fixed-Base Flight Simulator with a Dynamic Seat,* AIAA Modeling and Simulation Technologies Conference, 20-23 August 2007, AIAA-2007-6564

Heintzman, R. J. (1997). Determination of force cueing requirements for tactical combat flight training devices. Training Systems Product Group, Wright Patterson AFB, OH, ASC-TR-97-5001.

Heintzman, R.J. & Basinger, J.D. (1999). Development and validation of a method of evaluating the effectiveness of fighter aircraft simulation force cueing devices. SIMTEC Inc., Manassas, VA (DTIC ADA375328).

Heintzman, R. J., Middendorf, M, Basinger, J. D. (1999). Development and validation of a method of evaluating the effectiveness of fighter aircraft simulation force cueing devices. Training Systems Product Group, Wright Patterson AFB, OH, ASC-TR-2000-5001.

Jean, G. V. (2008). *Success of Simulation-Based Training is Tough to Measure.* National Defense, December 2008.

Lee, A. T. (2005). Flight Simulation: Virtual Environments in Aviation, Ashgate Publishing Company, Burlington, VT.

Martin, E. L., Waag, W. L. (1978). Contributions of platform motion to simulator training effectiveness: Study II – Aerobatics. Flying Training Division, Williams Air Force Base, AZ, AFHRL-TR-78-52.

Schreiber, B. T., Stock, W. A, & Bennett, W. Jr. (2006). Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Metric Development and Objectively Quantifying the Degree of Learning. (AFRL-HE-AZ-TR-2006-0015-Vol II). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Schreiber, B. T., Watz, E., Bennett, W. Jr. (2003). Development of a distributed mission training automated performance tracking system. Proceedings of the Behavioral Representations in Modeling and Simulation (BRIMS) Conference. Scottsdale, AZ.

Strachan, I. W. (2001). Motion cueing in the real world and in simulators – Principles and practice. In *Jane's Simulation and Training,* Surrey, UK: Jane's Information Group Limited.

Sutton, D., Skelton, M. & Holt, L. S. (2010). Application and Implementation of Dynamic Motion Seats. Intersevice/Industry Training, Simulation, and Education Conference (I/ITSEC) Proceedings. Orlando, FL.

Szczepanski, C., Leland, R. A. (2000). To move or not to move? A continuous question. Proceedings of American Institute of Aeronautics and Astronautics (AIAA), 0161.