Air Force Research Laboratory

**Distributed Mission Operations
Within-Simulator Training Effectiveness
Baseline Study: Metric Development and
Objectively Quantifying the Degree of Learning**

Brian T. Schreiber
William A. Stock

Lumir Research Institute
195 Bluff Avenue
Grayslake IL 60030

and

Winston Bennett, Jr.

Air Force Research Laboratory
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

Volume II

July 2006

Final Report for March 2002 to July 2005

**Human Effectiveness Directorate
Warfighter Readiness Research Division
6030 South Kent Street
Mesa, AZ 85212-6061**

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-AZ-TR-2006-0015-Vol II

This technical report has been reviewed and is approved for publication.

 // Signed //   **// Signed //**

WINSTON BENNETT JR.   **HERBERT H. BELL**
Project Scientist   **Technical Advisor**

 // Signed //

DANIEL R. WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* 03/31/2006	2. REPORT TYPE Final	3. DATES COVERED *(From - To)* March 2002 to July 2005

4. TITLE AND SUBTITLE

Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Metric Development and Objectively Quantifying the Degree of Learning

5a. CONTRACT NUMBER
F41624-97-D-5000

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)

Schreiber, Brian T.[1]; Stock, William A.[1]; & Bennett, Winston Jr.[2]

5d. PROJECT NUMBER
1123

5e. TASK NUMBER
AS

5f. WORK UNIT NUMBER
03

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

1. Lumir Research Institute
195 Bluff Ave
Grayslake, IL 60030

2. Air Force Research Laboratory
Human Effectiveness Directorate
6030 South Kent Street
Mesa, AZ 85212-6061

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFRL-HE-AZ-TR-2006-0015 – Vol II

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
This is the second volume of a five-volume report, Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study

14. ABSTRACT The current work reports only the objective data from AFRL-HE-AZ-TR-2006-0015, Volume I, Distributed Mission Operations Within-Simulator Training Effectiveness: Summary Report, but here we expand the reporting of objective data both in depth and breadth. We examined F-16 pilots participating in week-long Distributed Mission Operation (DMO) training exercises and compared extensive computer-collected data between beginning-of-week and end-of-week pilot performance on mirror-image scenarios. The DMO research environment in Mesa, AZ consisted of four high-fidelity F-16 simulators and one high-fidelity Airborne Warning and Control System simulator. Participating F-16 teams flew over 40 total scenarios according to a five-day syllabus, book-ended on Monday and Friday by mirror-image point defense air combat benchmark scenarios. Seven mission outcome measures were found to be significantly better on Friday than Monday: A 58.33% decrease in enemy strikers reaching their target, 38.10% greater distance from the base the F-16s disposed of the strikers, 54.77% fewer F-16 mortalities, 75.26% more enemy striker kills (before reaching base), 6.82% higher proportion of Viper Advanced Medium Range Air-to-Air Missile (AMRAAM) shots resulting in a kill, 51.60% lower proportion of enemy Alamo missile shots resulting in a kill, and a highly impressive 314.21% increase in an overall summary scoring scheme developed by subject matter experts. Significant trends were also found for a number of other metrics assessing skills. Of all the measures investigated in the current work, not a single offensive/defensive trade-off was observed, which significantly strengthens our conclusion that significant within-simulator learning took place.

15. SUBJECT TERMS
Distributed Mission Operations; DMO; Learning; MEC; Mission Essential Competencies; Metrics; Networked environments; Performance measurement; Pilot performance; Training effectiveness; Warfighter training

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Dr. Winston Bennett Jr
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED	UNLIMITED	42	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank

Table of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

The current work reports only the objective data from AFRL-HE-AZ-TR-2006-0015 Volume I, Distributed Mission Operations Within-Simulator Training Effectiveness: Summary Report, but here we expand the reporting of objective data both in depth and breadth. More specifically, in this report we discuss the importance of objective data, the development, and validation of the metrics, and report additional metrics and statistics. We examined F-16 pilots participating in week-long Distributed Mission Operations (DMO) training exercises and compared beginning-of-week to end-of-week performance on mirror-image scenarios. To evaluate performance, we collected extensive computer-based data of pilot performance (over 55 billion individual data points collected).

In conjunction with a computer-generated threat system and an instructor operator station, the DMO research environment in Mesa, AZ consisted of four high-fidelity F-16 simulators and one high-fidelity Airborne Warning and Control System simulator. From January 2002 to October 2004, participating F-16 teams flew over 40 total scenarios according to a five-day syllabus, book-ended on Monday and Friday by mirror-image point defense air combat benchmark scenarios. Seven mission outcome measures were found to be significantly better on Friday than Monday: A 58.33% decrease in enemy strikers reaching their target, 38.10% greater distance from the base the F-16s disposed of the strikers, 54.77% fewer F-16 mortalities, 75.26% more enemy striker kills (before reaching base), 6.82% higher proportion of Viper Advanced Medium Range Air-to-Air Missile (AMRAAM) shots resulting in a kill, 51.60% lower proportion of enemy Alamo missile shots resulting in a kill, and a highly impressive 314.21% increase in an overall summary scoring scheme developed by subject matter experts. Significant trends were also found for a number of other metrics assessing skills.

A large number of objective Mission Essential Competency (MEC)-based measures were defined, developed, and validated in the current work, but it is our recommendation that some skill measures be captured using other measurement means, such as expert observer ratings. In the objective data here, pilots performed better on almost every metric, including those that easily lend themselves to trade-offs (i.e., offensive and defensive metrics). The F-16 teams denied enemy strikers to base, killed more enemy aircraft, survived more frequently themselves, and did so while maintaining greater separation from the adversary (e.g., increased ranges in shots, F-poles, and decreased times in vulnerability zones such as Minimum Abort Range. Of all the measures investigated in the current work, not a single offensive/defensive trade-off was observed, which significantly strengthens our conclusion that significant within-simulator learning took place.

ACKNOWLEDGEMENTS

DISTRIBUTED MISSION OPERATIONS WITHIN-SIMULATOR TRAINING EFFECTIVENESS BASELINE STUDY: METRIC DEVELOPMENT AND OBJECTIVELY QUANTIFYING THE DEGREE OF LEARNING, VOLUME II

INTRODUCTION

Schreiber and Bennett (2006) reported a Distributed Mission Operations (DMO) training effectiveness study. That study represented a very large, comprehensive effort to evaluate DMO within-simulator training effectiveness, reporting numerous different data sources converging on the highly positive training effectiveness of the Mesa DMO environment. As such, that report's focus was to document the overall results stemming from the central hypotheses of each dataset and its scope, therefore, it prohibited reporting detailed results from any single dataset. The current work reports only the objective data from that study, but expands it both in depth and breadth. More specifically, in this report we discuss the importance of objective data, the development and validation of the metrics, and report additional metrics and statistics not suitable for AFRL-HE-AZ-TR-2006-015, Volume I, Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report.

In choosing indices for evaluating training improvement, scientists can pick from numerous approaches and methods (e.g., objective, expert observer ratings, opinions, surveys, mental models, decision making, etc.). As Bell and Waag (1998) point out, user acceptance is often necessary for a technology/training system to be seriously considered for routine or widespread use. Each additional potential data source and method can serve to address a critical facet of assessing learning. However, once acceptance is established (as we purport is the case with the Mesa DMO research site; Schreiber, Rowe, & Bennett, 2006), objective data arguably carries principal weighting among assessment methods. And, chief among objective data are the outcome metrics that measure warfighter performance exactly in the manner it will count during war—that is, kill ratios and mission objectives achieved. If powerful learning and transfer of training effects are discovered using these objective outcome metrics, from a warfighter's perspective all other effectiveness data types are relegated to secondary, more academic interest.

Additionally, automatically captured objective data affords powerful applied capabilities not easily equaled by other assessment methods. We can quantify the extent of DMO learning; we can more sensitively delineate differential learning performance among Mission Essential Competency (MEC) skills (Colegrove & Alliger, 2002); we can more accurately compare alternatives and their absolute degrees of effectiveness; we can, through standards (Schreiber, in press), cross-compare results across training institutions; we can potentially assess each and every warfighter in an exercise with a single assessment computer; we can calculate DMO return on investment; etc., etc. Furthermore, very unlike other psychological measurement techniques such as opinion data, surveys, mental models, or even expert observer ratings, objective data allows the warfighter more detailed feedback to diagnose performance. Finally, objective outcome data is difficult to argue with; no individual opinions, biases, or philosophies—only hard data that reports exactly what happened in terms of important combat-relevant metrics. Other data measurements are invaluable for rounding-out comprehensive effectiveness

evaluations, but the importance of reliable and valid objective data as central data cannot be overemphasized. However, challenges exist in obtaining the objective data.

One major challenge simply was developing a robust software tool that could reliably capture human performance data from multiplayer networks. Schreiber, Watz, Bennett, and Portrey (2003) reiterated an often stated need for robust objective measurement (e.g., Brecke & Miller, 1991; Kelly, 1998). Technological progress in several areas (especially the need for interoperability standards), however, have since matured and Schreiber et al. (2003) introduced a proof-of-concept software tool to capture data from a DMO network. Developmental performance measurement research at the Air Force Research Laboratory in Mesa, AZ resulted in this "Performance Effectiveness/Evaluation Tracking System" (PETS). PETS is a software tool that enables multi-platform, multi-level measurement ability at the individual and team level in a complex Distributed Interactive Simulation/High Level Architecture (DIS/HLA) environment. Installed at the Mesa research site, up to one million data points per minute are collected and organized into several formats differing in unit of analysis. Though a useable DMO software objective assessment tool obstacle appears recently overcome, another issue still looms.

The remaining major challenge is to properly identify standardized skills to be assessed and to define one or more metrics for a system like PETS to have meaningful data to capture. Identifying simple objective metrics for stand-alone simulation systems on simple tasks (e.g., emergency procedures) poses few complications compared to defining measures for air combat or for DMO training involving multiplayer networked environments. Defining objective assessments in complex tasks/environments presents much greater challenges. Fortunately, the MEC process has defined which skills constitute a proficient warfighter in combat, which are readily applicable to a realistic environment such as DMO. And, quite contrary to lower order emergency procedure-type skills exercised in standalone simulators, DMO can actually exercise the higher-order MEC skills (e.g., controls intercept geometry among several entities), which provides us potential opportunities to assess those skills in an ecologically valid environment. But the task of concretely developing/coding objective metrics for skills such as controls intercept geometry during many versus many scenarios must be undertaken. Once metrics are defined and validated, they can be programmed into the PETS human performance assessment tool and used for a great number of DMO research studies.

CURRENT WORK

The current work sought to fulfill the following specific objectives:

1. Define, develop, validate, and document air combat outcome metrics and objective skill metrics to be captured by the PETS assessment tool. Use the validated skill measures as standardized MEC-based assessment metrics for a DMO within-simulator effectiveness experiment and future DMO and live-fly studies.
2. Quantify the within-simulator learning benefits of five-day DMO training, report the comprehensive objective results here, and use the high-level results as a cornerstone database for the Volume I summary report (Schreiber & Bennett, 2006).

3. Analyze the objective data obtained above to explore potential moderators of DMO learning (e.g., flight experience).
4. Document more detailed results not suitable for reporting in AFRL-HE-AZ-TR-2006-0015-Vol I: Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report (Schreiber & Bennett , 2006).

METHODS

Overview

Subject matter expert (SME) interviews were used to identify critical behaviors for metric development and programming into the PETS system, a process which required approximately 24 months to complete. These metrics then served as the objective data source for the Volume I DMO within-simulator training effectiveness study (Schreiber & Bennett, 2006). In conducting said study, F-16 pilots arrived at the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division (AFRL/HEA) DMO training research facility in Mesa, AZ, for five days of training. The pilots received some simulator familiarization training and then were immediately "benchmarked," or "tested," on their pre-training point defense scenario performance. Post-training reassessment with those same pilots using mirror-image point defense scenario benchmarks occurred at the completion of five-day DMO training. The objective human performance metrics were collected throughout the five-day training. Observed performance between the pre- and post-test benchmark assessment sessions served as the basis for the within-simulator training effectiveness evaluation.

Metric Generation

To derive the measures, structured interviews were conducted with SMEs. A minimum of three SMEs were interviewed independently to each metric. We began air superiority measurement development with outcome metrics, which were defined by SMEs. We then asked SMEs to describe observable behaviors or events that constituted examples of good and poor performance. We identified skill metrics and the associated rule sets for a number of measures; the minor discrepancies found between SMEs during the independent interviews occurred only for a few metrics as the result of assumption differences and ideas overlooked. These differences were quickly resolved by bringing the SMEs together for concurrence. The air-to-air measure development interview process was undertaken before the air-to-air MECs were completed. As such, we later attempted to map the measures to the MEC skills using SMEs. We identified a number of the MEC air superiority skills with no associated measures (from the list of programmed metrics) and/or were deemed difficult to measure objectively at this time (e.g., "listens"). Therefore, a subjective assessment system was used in an attempt to capture many of those metrics (Schreiber, Gehr, & Bennett, 2006). We will discuss the objective metrics, both outcome and MEC-based skill metrics, in turn.

3

Objective Outcome Metrics

Strikers Reaching Base. For point defense scenarios, the most important goal is to deny enemy strikers from reaching their intended target. Given that the enemy strikers were carrying conventional bombs, the PETS system continuously tracked enemy strikers and "gave credit" to the strikers reaching base (and therefore an opportunity of employing weapons) if they flew within a two nautical mile radius (two-dimensional) of the intended target. Strikers reaching base is a dichotomous variable (yes/no) which carries utmost importance during war and point defense missions like those used in the current work. By far, it is therefore the single most important measurement among all five volumes of the study of the effectiveness evaluation. (Note: Strikers do not have to be killed to be "denied" reaching the base.)

Minimum Distance Achieved by Strikers. This is the closest distance to target achieved by strikers at any time during the scenario, reflecting a high-level look at the overall ease with which the F-16s were able to achieve mission success. Related to the metric above, this variable is a more sensitive continuous variable instead of a dichotomous one. If the enemy strikers flew unimpeded in the target area, this value would be very nearly or equal to zero.

Proportion of Vipers Killed. The proportion of all Vipers killed in the engagement. All scenarios contained four F-16s.

Proportion of (Valid) Enemy Strikers Killed. The proportion of enemy strikers killed during the benchmark scenarios before reaching the base (and therefore before having an opportunity to release weapons). All benchmarks contained two enemy strikers.

Proportion of all Threats Killed. The proportion of all enemy aircraft killed during the engagement. All benchmarks contained six enemy hostile fighters and two strikers.

Viper Missile Hit Proportions. Of all missiles fired, the proportion of which resulted in detonation on an enemy aircraft. This metric is reported by weapon type (AIM-9 and AMRAAM).

Threat Missile Hit Proportions. Of all threat missiles fired, the proportion of which resulted in detonation on an F-16. This metric is also reported by weapon type.

"Top Gun" Summary Outcome Scoring Scheme As a final outcome metric, F-16 SMEs defined a "Top Gun" scoring scheme as a single summary metric suitable for totaling performance on a single point defense scenario. These arbitrary scores were debated and settled upon after consideration of the point defense mission objectives and relative importance of each event. Point structure for the Top Gun summary metric is outlined as follows:

Enemy striker killed before reaching target: +450 points
Enemy striker killed after reaching target: +150 points
Enemy fighter (hostile) killed: +150 points
Fratricide: -900 points
Any other cause of F-16 mortality: -300 points

Objective Process/Skill Metrics

Process or skill oriented measures of performance are metrics assessing execution and typically correlate with outcome metrics. Frequently more sensitive to differences in skill, these process metrics often reveal significant changes in behavior and performance between individuals and/or teams when outcome metrics may be roughly equivalent. Results on each metric must frequently be taken into consideration with one or more other metrics to fully understand participant strategy biases or performance trade-offs. In developing each metric, we were limited to the information each simulator passed across the network according to DIS standards. Additional information aiding in assessing a skill had to be in the form of logic or configuration tables. DIS network traffic is fairly limited, primarily sending time, space, and positional information (TSPI) variables such as latitude, longitude, altitude, heading, pitch, roll, entity type, and airspeed. From these, however, we were still able to calculate a great number of other variables, such as aspect angles, closure rates, angles off tail, etc. Augmented with tables and algorithms housed on the PETS computer, we could then derive more complicated measures, such as those involving weapons envelopes. Nonetheless, since a great deal of relevant data remained resident within each simulator (e.g., symbology displaying results of weapons calculations), many skill measurements could not be developed at this time (i.e., the raw data is not available on the network). Standardizing more extensive interoperability data demands in the future could release more relevant information on the network that could then be used by the PETS system to automatically generate many more skill performance metrics.

Please note that the organization/mapping of the measures to a given MEC skill or supporting competency is still in progress. The organization of the measures reported on the following pages is the result of unanimous independent judgments by two SMEs highly familiar with both the MECs and the measures developed. With such a small sample of SMEs, the mappings listed herein should be considered preliminary.

MEC Skill: Weapons Employment

As long as meeting mission objectives and maintaining favorable kill ratios are not traded off, some generalities exist for effective weapons employment (e.g., launching a radar guided missile at higher altitude). Nine objective MEC weapons employment measures were developed:

Range at missile launch. At the time of pickle, the X and Y (i.e., 2D) and the X, Y, and Z (i.e., slant range) distances are computed. These distances are calculated for all missiles and are reported separately by weapon type.

Mach at Missile Launch. Mach at pickle is the velocity of the aircraft at the time a weapon is launched. This measure is calculated for all missiles and is reported separately by weapon type.

Loft Angle at Missile Launch. The loft angle at missile launch is the angle created between the aircraft's upward nose pitch and the imaginary level flight path at the time the

weapon is pickled. This measure is calculated for all missile shots taken outside of 10 nm and is reported separately by weapon type.

Altitude at missile launch. The altitude above Mean Sea Level (MSL) at the time of pickle. This measure is calculated for all missiles and is reported separately by weapon type.

Percentage Maximum at Launch. The percent maximum at firing is the reading of the caret within the F-16 dynamic launch zone (DLZ) at the time of pickle. That is, it is the percentage reading of maximum DLZ at the time of missile firing. This metric is calculated and reported for the AIM-9 and AMRAAM missiles. At the time of the current work, measurements in relation to R50 and R90 were not available (resident only within the simulator).

Escape-G at Launch. Escape-G at launch is a complex algorithm that takes into account ranges, closure velocity, aspect angles, altitudes, and weapon type to determine the exact degree of weapons engagement zone penetration. It is a transformed value of the DLZ reading, but measured in G-load units, and it can be thought of as the theoretical probability of a weapon intercepting its target at the time of weapon launch (i.e., an estimate of probability of kill). This value reports, at the precise time of pickle, the extent of G-load turn necessary for the targeted adversary to escape that weapon's fly-out (turning to either 180 or 0 aspect, whichever Escape-G value is lower) if the turn was to be initiated at the moment of pickle. This metric is derived from the same Fire Control Computer code that generates the DLZ in the F-16. This metric is calculated and reported for the AIM-9 and AMRAAM missiles. The Appendix contains a more detailed discussion on the Escape-G calculation and its theoretical background.

G-load at missile launch. The amount of G-forces on the aircraft at missile launch. This measure is calculated for all missiles and is reported separately by weapon type.

Distance of miss. The point of closest approach of the air-to-air munition employed. This metric, reported in feet, reports exactly how close a munition came to impacting on its intended target (a hit would therefore be zero). This measure is calculated for all missiles and is reported separately by weapon type.

Clear Avenue of Fire. The parameter of "Clear Avenue of Fire" (CAF) may be described as the degree to which the firing entity had a CAF from other friendlies to the intended target. That is, the degree to which another friendly may be at fratricide risk because of proximity to the intended threat. For the AIM-9, PETS measured the positional data for all aircraft in the engagement, drawing a line from the seeker head of the fired missile to the intended target, then calculating the angle of each friendly aircraft to that line during the entire fly-out of that missile. The nearest angular friendly to the fired missile during the entire fly-out (see Figure 1) represents the CAF. Rules for this measure ignored any friendly aircraft outside of five nautical miles. A similar, but different computation is made for the AMRAAM CAF.
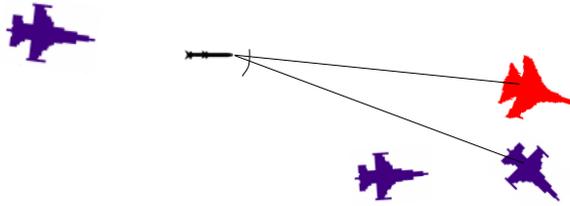
Figure 1 AIM-9 Clear Avenue of Fire

MEC Supporting Competency: Weapons Engagement Zone (WEZ) Management

These metrics are reported to provide indication as to what extent the friendlies kept themselves out of adversary WEZs. Generally speaking, disposing of all threats and achieving mission objectives while minimizing time in the Minimum Abort Range (MAR), Minimum Out Range (MOR), and Notch-Pole (N-pole) are desirable, as are maximizing such metrics as F-pole and A-pole.

MAR and MAR-1. For each scenario, the number of times a pilot allowed a hostile to fly within MAR is recorded, as well as the total time spent within MAR. These calculations are again made for MAR-1 nautical mile (i.e., 1 mile less than MAR values). These numbers are calculated according to the following rules:

1. All friendly aircraft ignored
2. All enemy strike aircraft (i.e., bombers) ignored
3. All enemy fighter aircraft position and weapon load tracked
4. Continuously calculates all hostile aspect angles
5. If aspect angle is > 120 degrees, then given the hostile's altitude, weapon type, and quadrant (refer to Figure 2), is range less than that of value in Table 1 below. If yes, friendly has allowed hostile to violate MAR.

For the time measures, one or more hostiles satisfying the above rules increments the timer (i.e., three simultaneous threats entering MAR does not triple the time), while the MAR count does count each hostile independently (i.e., three simultaneous threats entering MAR does triple the count).

MOR. For each scenario, the number of times a pilot allowed a hostile within MOR is recorded, as well as the total time spent within MOR. These numbers are calculated according to similar rules as the MAR metric.

N-pole. Similar to both MAR and MOR rule-set computations, the total time spent within N-pole and the number of violations of N-pole are computed.
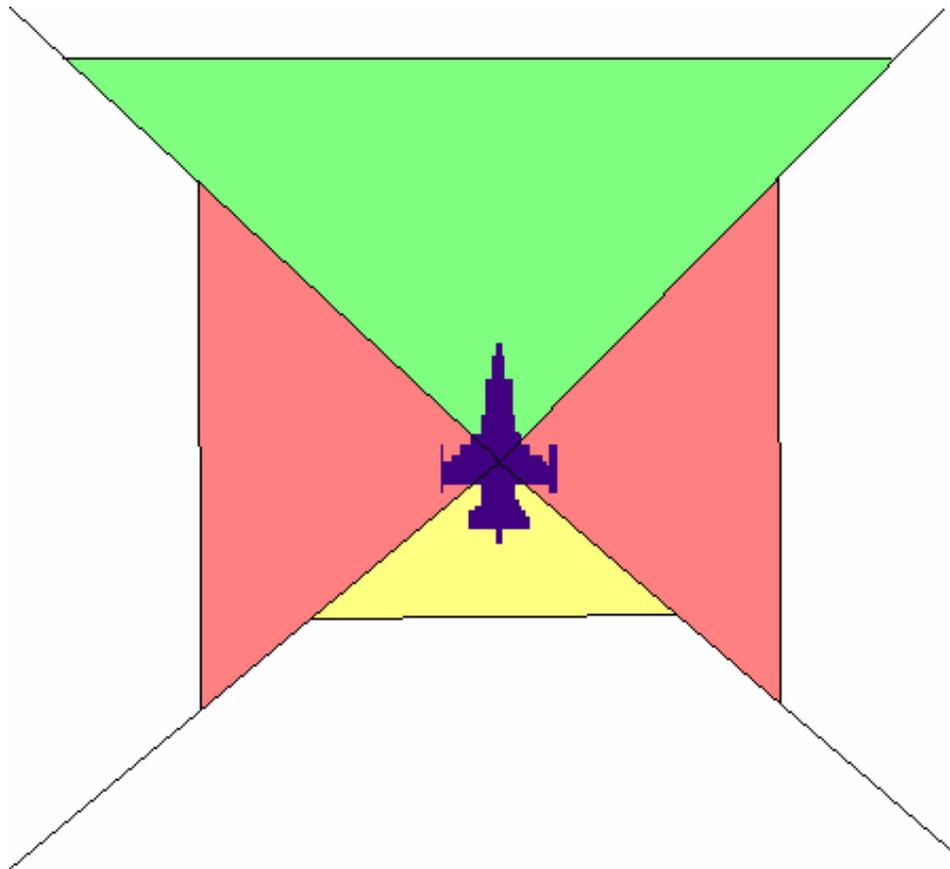
Figure 2 MAR, MAR-1, MOR, and N-pole heuristics

Table 1 Configuration table used by PETS to determine violations for MAR, MAR-1, MOR, and N-pole (note: cell values intentionally omitted).

	Adversary carrying AA-10A missile			Adversary carrying AA-10C missile		
	High Alt	Med Alt	Low Alt	High Alt	Med Alt	Low Alt
Front (MOR)						
Side (MOR)						
Rear (MOR)						
Front (MAR)						
Side (MAR)						
Rear (MAR)						
Front(MAR-1)						
Side (MAR-1)						
Rear (MAR-1)						
Front (N-pole)						
Side (N-pole)						
Rear (N-pole)						

F-Pole. Measured in feet, the F-pole range is the slant range from the firing entity to the target when the missile detonates. Ordinarily, fighter pilots define and use F-pole only for missiles that hit their target. Since the range preservation concept is important regardless of whether the missile hits the intended target, the F-pole metric is calculated for both hits and misses. For a missed shot, the PETS system captures and records F-pole the moment missile closure rate on the intended target begins to increase (i.e., increasing its distance from the target). Note: This measure is still undergoing validation.

A-Pole. A-pole is the distance from the launching aircraft to the target when a missile begins active guidance and is measured in feet.

Minimum 2D distance. The minimum 2D distance of a Viper to any threat fighter, measured in nautical miles.

MEC Skill: Maintains Formation

Wingman position. Wingman position is a measure of how often and for how long a wingman is out of formation. "In formation" is considered to be within three miles, 5,000 feet altitude and to be no more than ten degrees forward of the 3/9 line of the element lead. The measures are taken for Viper 2 from Viper 1 and from Viper 4 to Viper 3. The measures were also taken and separated for all time outside of 40nm to threats and time inside of 40 nm to threats.

Range between elements. This metric is a snapshot measurement taken at 30nm, 10nm, and 3nm to the nearest threat. The measures are slant range distances, measured in feet, between Vipers 1 and 2 to Vipers 3 and 4 (three total measurements taken).

Range within elements. This metric is a snapshot measurement taken at 30nm, 10nm, and 3nm to the nearest threat. The measures are slant range distances, measured in feet, between Vipers 1 and 2 and also between Vipers 3 and 4 (six total measurements taken).

Altitude between elements. This metric is a snapshot measurement taken at 30nm, 10nm, and 3nm to the nearest threat. The measures are distances, measured in feet, between Vipers 1 and 2 to Vipers 3 and 4 (three total measurements taken).

Altitude within elements. This metric is a snapshot measurement taken at 10nm and 3nm to the nearest threat. The measures are distances, measured in feet, between Vipers 1 and 2 and also between Vipers 3 and 4 (four total measurements taken).

MEC Skill: Controls Intercept Geometry

Altitude between V1/V2 and nearest threat. This metric is a snapshot measurement taken at 30nm, 10nm, and 3nm to the nearest threat. The measures are altitude separation distances, measured in feet, between the threat and Viper 1 or 2 (whoever is closer to threat; three total measurements taken). Because we report the measures in aggregate here over hundreds of engagements, the absolute values were used (instead of actual differences).

Altitude between V3/V4 and nearest threat. This metric is a snapshot measurement taken at 30nm, 10nm, and 3nm to the nearest threat. The measures are altitude separation distances, measured in feet, between the threat and Viper 3 or 4 (whoever is closer to threat; three total measurements taken). Because we report the measures in aggregate here over hundreds of engagements, the absolute values were used (instead of actual differences).

Range at first launch opportunity (FLO) Vipers 1 and 2. This metric, measured in nautical miles is a snapshot measurement taken at the first time the DLZ reads R50 and again at R90 for Vipers 1 or 2. (This measurement is taken regardless of threat designation or non-designation. The PETS system calculates all DLZs to all threats at all times to accurately capture this metric, regardless of what each individual Viper pilot chooses to display). Note: This measure is still in development and validation, and this measure is part of SCU-5 upgrades (vast majority of pilots in this study flew SCU-4).

Range at FLO Vipers 3 and 4. This metric, also measured in nautical miles is a snapshot measurement taken at the first time the DLZ reads R50 and again at R90 for Vipers 3 or 4. As with the above metric, this measurement is taken regardless of threat designation or non-designation. Note: This measure is still in development and validation, and this measure is part of SCU-5 upgrades (vast majority of pilots in this study flew SCU-4).

MEC Supporting Competency: Communication

Communication "step-overs" (frequency). This measures how many times two or more team members were attempting to communicate at the same time on the same frequency. This metric is captured for just the four Vipers as a team and again for the four Vipers + Airborne Warning and Control System (AWACS) as a team.

Communication "step-overs" (duration). Similar to the measure above, the communication duration measure calculates the average time of each communication "step-over." It is also calculated separately for the four Vipers as a team and the four Vipers + AWACS as a team.

Metric Validation

To ensure the accuracy and validity of each outcome and process measure, the following steps were undertaken (in chronological order):

1. Initial conceptual validity of each outcome and process metric was established through the structured SME interview process described in the "Metric Generation" section.
2. Each metric was transformed into C code and the rule sets were again presented to subject matter expertise before beta data collection.
3. To ensure we were capturing unusual scenario events and capturing measures correctly, simulator scenario set-ups were designed and flown with very specific, out of the ordinary trigger events in order to exercise all portions of software code (e.g., multiple

simultaneous shots at same entity, fratricides, entities killed as a result of flying into the ground trying to evade munitions, etc.).

4. Initial beta testing of the software was performed, collecting "test" data on operational pilots in the DMO environment. Software engineers identified and corrected bugs, if any.

5. Researchers and SMEs observed individual beta testing engagements in real-time and examined output files to confirm that the proper values of metrics were taken.

6. Outcome and shot-related metrics were provided as feedback to the beta testing operational F-16 pilot participants (zero inaccuracies reported).

7. Researchers plotted large sample distributions of each metric to ensure not only that all values did indeed fall within bounds for that metric, but also that the distribution properties observed adhered to expected values for that platform, missile, tactic, etc.

8. Trend data was checked across high/low experience demographics for improvements in the expected directions.

9. As a final validity check, a formal database was created for outcome metrics and shots. A research assistant, following a blind protocol, observed and manually recorded these same measurements for 163 scenarios. The human recorded data was then compared to PETS data of the same scenarios.

Participants

A portion of the following information is from General Method in Schreiber and Bennett (2006).

From January 1, 2002 to October 22, 2004, 76 fighter pilot teams participated in the overall DMO within-simulator training research study at the Mesa DMO site. An estimated 20% of the entire USAF F-16 worldwide population -- 384 pilots -- participated in the study. To participate in the training research, operational F-16 squadrons vied for posted vacant DMO training research weeks at the Mesa research site, readily volunteering for available training research opportunities. As such, participants in this study were not randomly sampled. Of the 76 teams under investigation for the overall study (Schreiber & Bennett, 2006), 53 teams produced useable objective data. Those teams that did not produce useable objective data were the result of either (a) not having at least two matched pairs of usable benchmark data, and/or (b) technical issues arising in the simulation environment that would systematically create biases in the objective data (e.g., temporary missile model change that dramatically increased missile probability of kill for several teams). Across the 272 pilots producing data useable for objective analyses, all but three were male, with a mean age of 33.1 years, 10.8 average years of military service, and a mean number of hours in an F-16 of 1,016 (Note: Between 3 and 5 of the 272 pilots did not provide information for one or more of the aforementioned demographic statistics and averages were computed based upon the remaining data).

DMO Training Facility

In conjunction with a computer-generated threat system and an instructor operator station (IOS), the DMO research environment in Mesa, AZ consisted of four high-fidelity F-16 simulators and one high-fidelity AWACS simulator. The F-16s, AWACS, and threat entities interoperated

according to DIS standards (IEEE Standard for Distributed Interactive Simulation - Application Protocols, 1995) version 4.02 or version 6.0.

The high-fidelity F-16 Block 30 simulators utilized 360 degree out-the-window visual displays with either SGI Onyx II Reality Monsters or PC Nova IIs running Aechelon runtime software. The visual system used high resolution photo-realistic databases of the Sonoran desert overlaid on terrain elevation data of the region. The hardware was very nearly identical to that found in the actual F-16, as was the software (Software Capabilities Upgrade version 4). Depending on the type of mission to be flown, F-16 weapon load-outs for missions consisted of differing combinations of the gun, the Air Intercept Missile (AIM-9), the Advanced Medium Range Air-to-Air Missile (AMRAAM), and/or the Mk-82 and Mk-84 general purpose bombs. A high-fidelity Solipsys version 6 AWACS sensor simulation was also used to provide a more realistic environment.

The Automated Threat Engagement System (ATES) generated all adversaries. A computerized, real-time threat generation system, ATES operates on standard DIS networks, providing air-to-air, air-to-ground, and surface-to-air threats. The ATES incorporates aerodynamic modeling, atmospheric models, radar models, infrared models, and data parameter tables for thrust, drag, lift, etc. For the current work, threat air models were the MiG-29, MiG-27/23, and Su-27 loaded with the AA-8, AA-10a, and AA-10c air-to-air missiles. Ground threats included the SA-2, SA-6, and SA-8, and AAA. Threat aircraft performed maneuvers and/or scripted flight paths while reacting to the F-16's maneuvers and weapons.

Throughout the majority of the data collection time period, the debrief facility included five 50-inch plasma screens -- one for a God's eye view and one dedicated for each of the four F-16s. Each of the F-16 plasma screens presented four avionic displays from the F-16. The time synchronized replay included all communications and could be paused, fast-forwarded, or rewound according to the lead pilot's desired use of the allotted debrief time.

As a training research installation striving to continually integrate and evaluate new training technologies, the DMO site at Mesa undergoes occasional upgrades to its simulation systems. Therefore, the DMO simulation environment was *not* constant for all participants in this study. Some examples of upgrades/changes to the environment during the 33-month data collection period included (but is not limited to):

- Upgrading the visual databases in cockpits #3 and #4 to use the same photospecific database used in cockpits #1 and #2,
- upgrading to eight visual channels,
- upgrading the radios,
- installing SCU-5 Situation Awareness Data Link (SADL) software,
- installing new ALQ-213 radar warning/electronic countermeasure panels and 5100 power PC boards,
- adding smoke trails to missile fly-outs,
- upgrading the brief/debrief facility with Portable Flight Planning Software version 3.2, and
- a sixth 50-inch plasma debrief display for AWACS.

Under most circumstances changing the apparatus during the course of a scientific study threatens the study's conclusions. However, for the current work, we viewed these changes in the DMO environment as highly desirable. Further explained, as a system of integrated technologies, _all_ _DMO_ _environments_ _will_ _change_ and be constantly upgraded at every field location. By doing similarly in our experimental environment we more closely replicate the actual systems to which we aim to generalize. Furthermore, we argue that significant learning effects must be found in light of the additional error variance associated with updates/changes to the environment, because of the fact that DMO environments will undoubtedly undergo change. If a training effect is not found under these changing conditions, justification for DMO training does not exist.

Training Research Syllabi/Training Research Week.

Table 2 shows a general timeline for each participating team. Participants arrived early Monday morning for five days of DMO participation. Upon arrival, participants were first given an inbrief on the objectives and procedures of DMO and the simulators, a tour of the facilities, and then given a research administrative session where they completed a demographic form, were assigned anonymous barcode identification numbers, and finally took the first Pathfinder exercise-- an electronic assessment used to capture the knowledge structures of novice and expert pilots (Schreiber, DiSalvo, Stock, & Bennett, 2006).

Table 2 Participant General Timeline.

Session#	1	2	3	4	5	6	7	8	9
Day/time	Mon AM	Mon PM	Tues AM	Tues PM	Wed AM	Wed PM	Thur AM	Thur PM	Fri AM
Activity	Mesa Inbrief Admin Pathfinder Pilot Brief Fly Fam Pilot Debrief	Pilot Brief Fly 3 Benchs+ Pilot Debrief Feedback Survey	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 3 Benchs+ Pilot Debrief Feedback Survey Reaction Survey Pathfinder Outbrief

Pilots participated in one of four very similar syllabi, each syllabus consisting of nine 3.5 hour sessions, beginning with session one on Monday morning and ending with session nine on Friday morning. There were two sessions each day of the five-day training week, save Friday's single session. Each session entailed a one hour briefing, an hour of flying multiple engagements of the same mission genre, and an hour and a half debriefing. The syllabi scenarios could be either offensive or defensive, but were all four F-16s versus X number of threats. Scenarios were designed with trigger events and situations to specifically train MEC skills (Symons, France, Bell, & Bennett, 2006). These syllabi were developed with traditional

13

methods using full mission rehearsal scenarios across a spectrum of probable air-to-air missions and threats while increasing the complexity of the missions as the training research week progressed.

After completing the administrative tasks early Monday morning, each syllabus began with a familiarization session (session one) late Monday morning to orient pilots to DMO simulator environment specifics, such as visual ID characteristics and any switchology differences due to F-16 block number or F-16 mission software. The pilots required surprisingly little familiarity training. The hour allotted turned out to be more than enough familiarity time, as the high fidelity simulator layout and underlying simulation models closely resembled the actual aircraft and pilots very quickly became comfortable with DMO simulator operation. Since the pilots readily and easily adapted to the simulation environment during the familiarization period, performance increases observed throughout the course of the subsequent sessions should be the result of learning/honing their skills and not learning "sim-isms" or other DMO idiosyncrasies.

Session two on Monday afternoon began with benchmarks (i.e., a "pre-test") used to measure pre-training performance. The training week ended with the "post-test" training benchmark session nine on Friday morning. The benchmark sessions consisted of flying three point defense engagements (see Figure 3). All benchmark point defense scenarios pitted the four participant F-16s and their AWACS controller against eight threats (six hostiles and two strikers) at a distance greater than 40 nautical miles. During all benchmark scenarios, AWACS informed the F-16s (at long range to the threats) that there were six entities and that all six were already identified as hostile, thereby allowing the F-16s to shoot beyond visual range at those six entities. Regarding the two strikers, the AWACS operator could not "see" below 10,000 feet--the altitude under which the enemy strikers flew during all benchmarks. Therefore, the onus fell upon the F-16s to find any entities below 10,000 feet with their onboard radars and visually identify them before employing ordnance.

All benchmarks were designed to be equally complex according to the absolute complexity scoring scheme outlined by Denning, Bennett, and Crane (2002). Seven-point defense benchmark scenarios were developed, and the complexity analysis revealed that all benchmarks were indeed equally complex. Pilots flew in the same flight/cockpit assignment on Monday and Friday. Unbeknown to the pilots, for the Friday benchmarks, pilots flew the mirror image of the three benchmarks that were flown on Monday. Strict data collection rules governed all benchmarks in order to maintain a realistic combat environment—i.e., no freezing or reloading entities, fuel always on, no reincarnating entities, no inserting new entities, real-time kill removal for all entities, no intervention/assistance from IOS operators, etc. Benchmarks terminated under one the following conditions: All F-16s dead, all air adversaries dead, enemy strikers reached their target, or 13 minutes elapsed time. During the course of the study, the vast majority of benchmarks terminated under one of the first three conditions.

The participants' overriding goal for the point defense benchmark scenario was to prevent the enemy strikers/bombers from reaching the base – success being striker denial or kill. The second and third most important goals are to minimize friendly mortalities and maximize the adversary kills. The point defense benchmark scenarios were selected for examination in the present study as pre- and post-test assessments because: (1) point defense scenarios have very clear goals and

14

measures of success, (2) all the benchmark engagements have equivalent levels of complexity, (3) three benchmark scenarios occur at the beginning and the end of the week-long DMO syllabus, (4) the same pilots in the same cockpit assignments perform the mirror-image benchmark scenarios at the beginning and the end of the week (unknown to them), and (5) the benchmarks were flown under real-time kill removal and strict data collection rules.
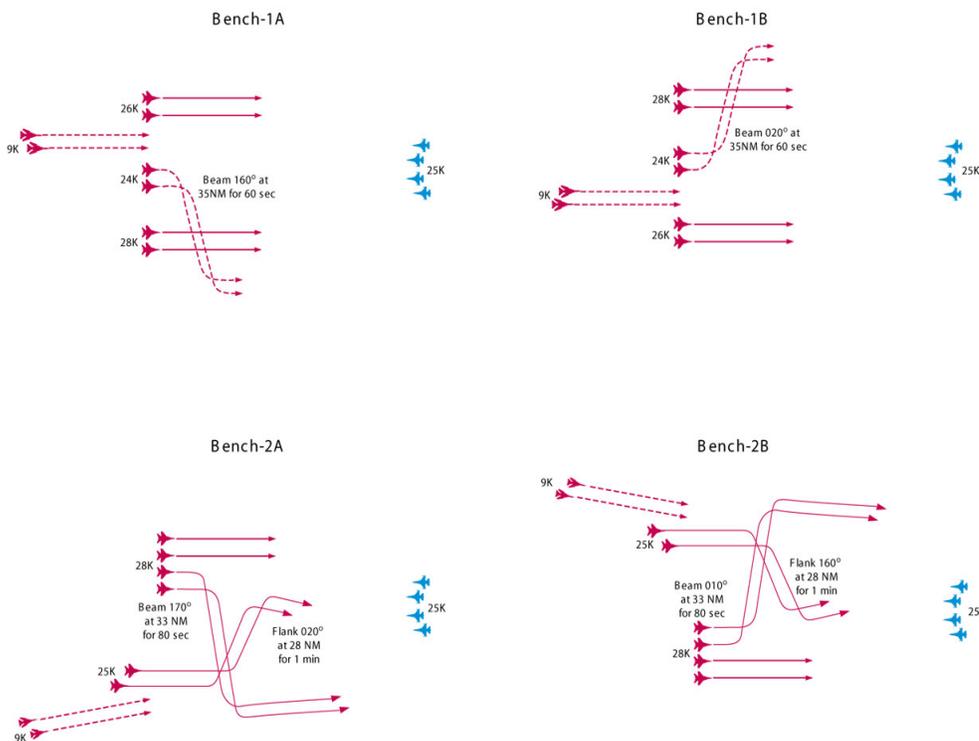
Figure 3 Example mirror-image point defense benchmark scenarios used for the pre- and post-test

The MEC-based building-block training began immediately after the benchmarks (with the remaining time during session two) on Monday afternoon and continued through the course of the week. Participating teams were exposed to four to eight full engagements per session. While these training sessions emphasized Defensive Counter Air scenarios (DCA), pilots also flew some Offensive Counter Air (OCA) and air-to-ground missions. Usually, participating teams experienced about 35 training engagements between the Monday and Friday benchmarks, providing an intensive training curriculum. The building block training sessions progressed in complexity by increasing the number of threat aircraft, the type of threat aircraft, the threat aircraft reactivity/maneuver, and/or an increase in the vulnerability time.

Either after the last session on Thursday or on Friday morning, pilots took the second Pathfinder exercise and were given a DMO reaction rating form. The DMO rating form is a rating scale survey that pilots use to rate their DMO training experience. After the last session on Monday and Friday, the teams were also given a self-report feedback form with open-ended questions asking if they felt their objectives had been met and what facilitated or hindered their performance. Finally, before departure, teams were given a performance outbrief after their last

set of benchmarks. This outbrief consisted of graphs for a number of the objective measures, showing the team's observed performance.

RESULTS

Data

This report omits certain results. At the time of press, intentions were to generate an additional report with additional data suitable for restricted distribution channels.

In the results section that follows, t-test degrees of freedom vary across the variables tested. A primary cause of this variation is the lack of an observation on the variable in question. Because Monday and Friday for each variable were compared using dependent t-tests and because these tests require responses to be present for both days, the absence of an observation on either Monday or Friday causes the loss of a decrease in the degrees of freedom for the test.

Metric Validation

As previously discussed, metric validation was a nine-step process. Progression from each step to the next was undertaken only when complete confidence on the prior developmental validation step was achieved. As such, here in the results section we report only the final step, comparing human observed data to automatically captured data for those metrics suitable for human counting. A research assistant observed 163 benchmark scenarios, manually recording (a) strikers on target, (b) F-16 mortalities, (c) total threat mortalities, and (d) missiles fired. The correlation between each of these four human observed metrics and the corresponding PETS captured metric was .75, .98, .92, and .94, respectively. Though all these correlations were statistically significant, we originally anticipated even higher correlations. Upon further investigation, explanations for the small differences became readily apparent. For the bombers to target, the debrief system's God's-eye view was very difficult for the research assistant to use in determining if strikers reached a 2nm radius of the target, as the overhead view was typically depicted in 10nm square grids with no circular rings to aid judgments. As far as the enemy mortalities, there were a few instances where a threat was being chased, but was just out of range of an AIM-9, and the threat eventually flew into a mountainside or the ground. Additional logic code within PETS gave kill credit to the chasing F-16, whereas the research assistant did not. A small discrepancy in shot correlation was also expected, as we have found in other research that even F-16 subject matter experts fail to count shots perfectly over a large number of graded engagements (Krusmark, Schreiber, & Bennett, 2004). This final validation step and subsequent investigation resulted in greater confidence of the PETS-derived outcome measures than the more error-prone manually recorded outcome metrics.

Outcome Metrics.

Of the 76 teams reported in AFRL-HE-AZ-TR-2006-0015-Vol I, Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report, 53 teams (272 pilots) produced usable data for some or all of the objective data analyses. Table 3 contains the summary results for all outcome metrics collected on benchmarks.   A t-test procedure was

performed for each metric. Significant Monday to Friday differences were observed on 8 of the 10 outcome metrics, and *all* of those effects were in the expected direction (i.e., improved performance). Number of strikers reaching base reduced 58.33% from Monday to Friday [$t(52) = 4.70$, $p < .01$], while the closest average distance they came to target increased 38.10% [$t(52) = -2.21$, $p < .04$]. F-16 mortalities reduced 54.77% ($t(52) = 4.76$, $p < .01$), while total threats killed and enemy strikers killed (before reaching the defended base) increased 9.20% and 75.26% [$t(52) = -4.84$, $p < .01$ and $t(52) = -6.34$, $p < .01$]. The proportion of F-16 AMRAAM missiles resulting in a kill increased 6.82% [$t(51) = -2.23$, $p < .03$], while the proportion of threat Alamo missiles resulting in a kill decreased 51.60% [$t(49) = 5.35$, $p < .01$]. Combining many of the outcome measures, it comes as no surprise then that we also found a significant increase in the "Top Gun" summary scoring scheme—an impressively large 314.21% increase [$t(52) = -5.62$, $p < .01$].

Table 3 Summary results for all outcome metrics.

Metric Name	Mon vs. Fri	
# of enemy strikers reaching target	Decreased by 58.33%	
Closest distance achieved in above	Increased by 38.10%	
# of Viper mortalities	Decreased by 54.77%	
# of enemy strikers killed (before base)	Increased by 75.26%	
Total # of enemy threats killed	Increased by 9.20%	
"Top Gun" summary scoring scheme	Increased by 314.21%	
	BVR Missiles	*Heat-seeking*
Prop. Viper missiles resulting in a kill	Increased 6.82%	NS
Prop. Threat missiles resulting in a kill	Decreased 51.60%	NS
Cells in italics represent a statistically significant Monday to Friday change, p<.05		

To determine if any demographic variables moderated the learning, we specifically examined the following for the lead pilot of the session 2 and 9 benchmarks (taken from the demographic questionnaire):

1.   Flight Qualification (wingman, 2-ship lead, 4-ship lead, Mission Commander, and Instructor Pilot).
2.   Weapons Instructor Course (WIC) Graduate (yes/no)
3.   Total number of DMO Simulator Exercises (e.g., Shaw Mission Training Center [MTC])
4.   Total number of Live-Fly Exercises (e.g., blue/red/green/Maple Flag)

As we felt it produced the best summary of all the important outcome metrics by combining both offensive and defensive measures, we examined each of the above four demographics against the "Top Gun" scores for moderating effects. Since Flight Qualification and WIC Graduate are categorical variables, we ran Analysis of Variance (ANOVAs) for these and examined the interaction between these demographics and the Monday/Friday change in Top Gun scores. Pearson correlations were calculated for the continuous variables of Number of Live-Fly Exercises and Number of DMO Simulator Exercises. We found no significant effect for Flight Qualification, $F(4,39) = 2.32$, $p = .43$, WIC Graduate, $F(1,42) = 2.21$, $p = .14$, or Number of DMO Simulator Exercises, $r = .07$, $p = .66$, but we did find a significant correlation with Number of Live-Fly Exercises, $r = .34$, $p < .03$.

MEC Skill "Weapons Employment"

Table 4 contains the summary results for all MEC skill "Weapons Employment" assessment metrics. A t-test procedure was performed for each metric. Significant Monday to Friday differences were observed for some, but not all measures. *All* of the significant effects were in the expected direction (i.e., improved performance). Comparing Monday to Friday, the F-16 pilots, on average, pickled the AMRAAM at significantly increased 2D ranges [10.30% longer; $t(52) = -3.70$, $p < .01$], slant ranges [10.31% longer; $t(52) = -3.71$, $p < .01$], mach [5.28% faster; $t(52) = -4.44$, $p < .01$], loft angle [14.80% higher; $t(52) = -2.66$, $p < .01$], and altitudes [7.97% higher; $t(52) = -4.05$, $p < .01$] on Friday, but only one of these was significant for the AIM-9 [mach at pickle increased 11.02%; $t(46) = -2.39$, $p < .02$].

Table 4 Summary results for all MEC skill "Weapons Employment" assessment metrics.

Metric Name	Mon vs. Fri	
	AMRAAM	AIM-9
Range at pickle (2D)	Increased 10.30%	NS
Range at pickle (slant)	Increased 10.31%	NS
Mach at pickle	Increased 5.28%	Increased 11.02%
Altitude at pickle	Increased 7.97%	NS
Loft angle at pickle	Increased 14.80%	NS
G-loading at pickle	NS	NS
Percent of DLZ maximum at pickle	NS	NS
Escape-G at pickle	NS	NS
Distance of miss	NS	NS
Clear Ave of Fire	NS	NS
Cells in italics represent a statistically significant Monday to Friday change, p<.05		

MEC Supporting Competency "Weapons Engagement Zone Management"

Table 5 contains the summary results for all MEC supporting competency "WEZ Management" assessment metrics. A t-test procedure was performed for each metric. Significant Monday to Friday differences were observed for over half of the WEZ management measures, and all those were in the expected direction. Comparing Monday to Friday, the F-16 pilots, on average allowed hostiles into MOR, MAR, MAR-1, and N-pole for significantly less time (respectively, decreases of 14.15%, 55.20%, 57.90%, and 60.33% with t-values (52) of 2.51, 4.00, 3.88, and 4.99, and associated p-values less than .02, .01, .01, and .01). The number of times the F-16 pilots allowed hostiles into MAR [-39.92%, $t(52) = 3.53$, $p < .01$] and MAR-1 [-44.56%, $t(52) = 3.45$, $p < .01$) were also significantly reduced. The minimum range any F-16 pilot came to any hostile during a given benchmark increased by 44.99% [$t(52) = -4.85$, $p < .01$). These increased ranges from the threats reveals itself in the weapons fly-out tactical behavior as well, where A-pole ranges were 14.35% longer [$t(34) = -3.15$, $p < .01$] and F-poles were 8.12% longer [$t(52) = -2.60$, $p < .01$].

Table 5 Summary results for all MEC Supporting Competency "WEZ Management" assessment metrics

Metric Name	Mon vs. Fri
Hostiles in MAR (count)	Decreased by 39.92%
Hostiles in MAR (time)	Decreased by 55.20%
Hostiles in MAR-1 (count)	Decreased by 44.56%
Hostiles in MAR-1 (time)	Decreased by 57.90%
Hostiles in MOR (count)	NS
Hostiles in MOR (time)	Decreased by 14.15%
Hostiles in N-pole (count)	++
Hostiles in N-pole (time)	Decreased by 60.33%
F-pole (hits and misses; AMRAAMs)	Increased by 8.12%
A-pole (AMRAAMs)	Increased 14.35%
Minimum 2D range to hostile	Increased by 44.99%
Cells in italics represent a statistically significant Monday to Friday change, $p < .05$. A cell with a ++ denotes that the value was not output at time of going to press.	

MEC Skill "Maintains Formation"

Table 6 contains the summary results for all MEC skill "Maintains Formation" assessment metrics. A t-test procedure was performed for each metric. Significant Monday to Friday differences were observed on just 3 of the 24 metrics.

Table 6 Summary results for all MEC skill "Maintains Formation" assessment metrics

Metric Name	Mon vs. Fri		
# times V2 violated wingman position (>40nm to threats)	NS		
Prop. time V2 spent in wing. pos. violation (>40nm to threats)	Increased 15.46%		
# times V4 violated wingman position (>40nm to threats)	NS		
Prop. time V4 spent in wing. pos. violation (>40nm to threats)	Decreased 12.24%		
# times V2 violated wingman position (<40nm to threats)	NS		
Prop. time V2 spent in wing. pos. violation (<40nm to threats)	NS		
# times V4 violated wingman position (<40nm to threats)	NS		
Prop. time V4 spent in wing. pos. violation (<40nm to threats)	NS		
	30nm	10nm	3nm
Range btw elements (1/2 to 3/4)	NS	Decreased 10.00%	NS
Range within element 1/2	NS	NS	NS
Range within element 3/4	NS	NS	NS
Alt btw elements (1/2 to 3/4)	NS	NS	NS
Alt within element 1/2		NS	NS
Alt within element 3/4		NS	NS
Cells in italics represent a statistically significant Monday to Friday change, p<.05			

MEC Skill "Controls Intercept Geometry"

Table 7 contains the summary results for all MEC skill "Controls Intercept Geometry" assessment metrics. A t-test procedure was performed for each metric. Significant Monday to Friday differences were observed for the altitude between V1/V2 and nearest threat at 30 nm [increased by 32.76%; t(52)= -4.26, p<.01], and the altitude between V3/V4 at 30nm [increased by 22.48%; t(52) = -3.66, p < .01], 10nm [increased 18.23%; t(51) = -1.99, p < .05], and 3nm [increased 36.24%; t(25) = -2.27, p < .03]. As the pilots in this study flew SCU version 4 (and the fact that creating measures for R50/R90 from SCU 5 were still in development), outputs for R50/R90 were not obtained for the current work.

Table 7 Summary results for all MEC skill "Controls Intercept Geometry" assessment metrics

Metric Name	Mon vs. Fri		
	30nm	10nm	3nm
Altitude btw V1/V2 & nearest threat	Increased 32.76%	NS	NS
Altitude btw V3/V4 & nearest threat	Increased 22.48%	Increased 18.23%	Increased 36.24%
	R90		R50
Range @ FLO for V1/V2	++		++
Range @ FLO for V3/V4	++		++
Cells in italics represent a statistically significant Monday to Friday change, p < .05. *Cells with a "++" symbol were not calculated for the current work.*			

Table 8 contains the results of communication use inside of 40nm to the threats. These communication metrics only report the frequency and duration of communication "step-overs" by measuring the unique instances participants "push to talk" on the radio at the same time on the same frequency. A t-test procedure was performed for each metric. Significant Monday to Friday differences were observed for three of the four measures, all declines. Frequencies of step-overs were 34.61% less for the entire team [t(52) = 4.40, p < .01] and 16.33% less for just the F-16 four-ship [t(52) = 2.66, p < .01]. Durations of step-overs were not significantly different for the whole team, while durations of step-overs were significantly less (-11.92%) for just the F-16 four-ship [t(52) = 2.90, p < .01].

Table 8 Radio communication use.

Metric Name	Mon vs. Fri
"Step-over" frequency within Viper flight	Decreased 16.33%
"Step-over" frequency among team	Decreased 34.61%
"Step-over" duration within Viper flight	Decreased 11.92%
"Step-over" duration among team	NS
Cells in italics represent a statistically significant Monday to Friday change, p<.05	

DISCUSSION

Satisfying our first objective – defining, developing, validating, and documenting air superiority MEC-based metrics – proved overall slightly more challenging than originally anticipated. During the SME interview process, we defined a number of skill measures rather easily, but found the translation difficulty into software code varied significantly with each metric. Many weapons launch metrics, for example, were quite easy to capture off the DMO network and could be coded, tested, and validated within a few weeks. However, complete and accurate DMO network data became an issue that prohibited us from developing/assessing some measures. We say "given complete and accurate" due to (a) current standardized DIS network data are not terribly comprehensive and much simulator data are not passed on the network, halting measurement development entirely for some skill metrics, and (b) the fact that some DMO network data would occasionally be missing (e.g., who fired a missile). For these latter "occasional missing data" instances, additional time and logic was required to accurately capture the skill metric. For other, more complicated measures (e.g., Escape-G), many months of coding, testing, and validation were required. Finally and unfortunately, we discovered that many of the MEC air superiority skills simply did not lend themselves to automated objective assessment (e.g., "listens"). As a result, despite our continued automated objective metric development efforts, it is our recommendation that some skill measures be captured using objective techniques like those described herein and some skill measures be captured via other measurement means, such as expert observer ratings.

Identifying and defining mission outcome metrics was very straightforward in the SME interviews, but coding them and obtaining the measures reliably and accurately was altogether more complicated. As one example, if a friendly chased an adversary but was not quite within range for missile employment, the adversary, on occasion, might have flown into a mountainside in its attempt to escape; kill credit needed to be given to the chaser. As another example, if two friendlies each fired a missile and both detonated on the same adversary at roughly the same time, only one "kill" can be registered and therefore only one friendly can receive kill credit. In both these examples (unforeseen before development began), we wrote significant extra logic code to simply track shots and lives lost in order to assign kill credit correctly.

For outcome and skill measures, completing metric development, coding, and validation required approximately two years, but the results were worth the effort. All metrics underwent a multi-step validation effort and the most critical outcome metrics were subjected to a final manual observation validation step where correspondence/validation was solidified. Consequently, millions of individual data points now can be captured automatically and reliably on MEC-based metrics. These MEC-based metrics are invaluable for DMO studies.

The first such DMO study (and our second objective, which is documented in this report) was quantifying the within-simulator learning benefits of DMO training. Previously, DMO assessments largely consisted of expert observer ratings, which showed effects, but often were not very sensitive in differentiating skills. We attribute this to a number of factors, two primary ones including observer anchoring and lack of SME observer measurement sensitivity. The objective metrics here were much more sensitive than what we have experienced with SME ratings (Krusmark, et al., 2004; Schreiber, Gehr, & Bennett, 2006), revealing a number of highly

significant effects in many different areas. Most importantly, strikers to target and friendly mortalities by Friday changed an astounding -58% and -55%, respectively. Indeed, during informal mission observation, we frequently noticed that the F-16 pilots were so preoccupied with hostiles on Monday benchmarks that they infrequently engaged (or, often, did not even sample with radar!) the enemy strikers. On Friday benchmarks, enemy strikers were much more easily disposed of and done so at sufficient range from the point being defended (closest distance achieved by strikers on Friday increased by 38%). These highly significant results lay to rest any doubts that DMO training yields substantial within-simulator learning, at least for F-16 air superiority missions.

Participants were unaware of most measures used for evaluation. That is, the pilots only knew we were assessing their performance in a general sense. Sometimes, participants who know they are being evaluated may maximize their performance on one assessed dimension at the sacrifice of another task dimension so as to achieve a goal criterion (e.g., speed/accuracy trade-off). In the data here, pilots performed better on almost every metric, including those that easily lend themselves to trade-offs (i.e., offensive and defensive metrics). The F-16 teams denied enemy strikers to base, killed more enemy aircraft, survived more frequently themselves, and did so while maintaining greater separation from the adversary (e.g., increased ranges in shots, F-poles, A-poles, and decreased times in vulnerability zones such as MAR). Of all the measures investigated in the current work, not a single offensive/defensive trade-off was observed. The fact that the offensive and defensive measures improved simultaneously and significantly over the course of the week deeply strengthens our conclusion that significant within-simulator learning took place.

After SME review, explanation for the large proportion of non-significant findings for the Maintains Formation metrics (Table 6) became very straightforward. The metrics in the bottom half of the table (i.e., those snapshots taken at 30, 10, and 3nm) are situation-specific and valuable for mission debrief. However, in the aggregate, general rule-of-thumb performance generalizations are not easily made. Therefore, for future application we recommend that these measures be used for mission-by-mission training feedback and not as part of future data aggregated effectiveness evaluations. Wingman formation (i.e., those measures in the top half of Table 6) is a more fundamental skill. That is, wingman formation is a lower order skill, and since DMO scenarios like those used in the current work tend to exercise higher order skills, a SME would not expect numerous significant learning effects on those measures.

Future Directions.

If these results transferred entirely to combat, the capability of force gained from a single week of DMO training easily justifies the expenditures paid for DMO simulation environments. But, that would, of course, be a potentially error-prone extrapolation. Additional research should be undertaken to understand how quickly the gains decay and to what extent the gains transfer to live-fly range activities representative of actual combat (initial efforts for both of these studies are being undertaken). Additionally, attempting to generalize results concerning DMO from the current work is limited by the use of just F-16 point defense missions and a non-random participant pool. With additional research addressing the above-mentioned areas, we can better understand the different facets of DMO training benefits.

During the course of the current work, it became evident that measuring a particular MEC skill would require multiple measures. Ideally, we would have liked to use the multiple measures for each MEC skill to develop a composite assessment for that skill. One simple approach would have been to build multiple regression models for each skill. Though we had numerous predictor variables for this, we did not have any suitable global assessments of each MEC skill to use as a dependent variable to build the models. One approach in developing a suitable dependent variable may be to use new (and blind) SME ratings for just those MEC skills across a sufficient sample of engagements. Then an effort could be undertaken to build and validate composite assessment models using different datasets.

Our goal is to build both summary level measures (i.e., scenario summary measure) and real-time measures (i.e., as the pilot is flying) to assess all of the MEC skills. The summary level measures prove useful in overall assessments of competency and for aggregating data across scenarios, pilots, and teams. Real-time measures as the pilot is flying would serve as useful diagnostic tools to an instructor/evaluator. As previously mentioned, we see both objective and subjective tools as necessary to assess all the MEC skills. Once this is accomplished, those measures could then be used as part of an adaptive and continuous learning system. That is, warfighters could continually be monitored on their performance—delineated by the various MEC skills—and real-time deficiencies on certain skills would be identified. These deficiencies could then be specifically targeted by MEC-based scenarios with trigger events specifically tailored to train those skills.

REFERENCES

Bell, H. H, & Waag, W. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. *The International Journal of Aviation Psychology, 8(3), 223-242.*

Brecke, F.H., & Miller, D.C. (1991). *Aircrew performance measurement in the air combat maneuvering domain: A critical review of the literature* (Rep. No. AL-TR-1991-0042, AD B158 404). Williams Air Force Base, AZ: Armstrong Laboratory, Aircrew Training Research Division.

Colegrove, C. M. & Alliger, G. M. (2002, April). *Mission Essential Competencies: Defining combat mission readiness in a novel way*. Paper presented at the NATO RTO Studies, Analysis and Simulation (SAS) Panel Symposium. Brussels, Belgium.

Denning, T., Bennett, W. Jr., & Crane, P. M. (2002). Mission complexity scoring in Distributed Mission Training. In *2002 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Kelly, M. J. (1988). Performance measurement during simulated air-to-air combat. *Human Factors*, 30, 495-506.

Krusmark, M., Schreiber, B. T., & Bennett, W., Jr. (2004). The effectiveness of a traditional gradesheet for measuring air combat team performance in simulated Distributed Mission Operations. (AFRL-HE-AZ-TR-2004-0090, ADA428119). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schreiber, B. T. (in press). Transforming training: A perspective on the need and payoffs from common standards. *Journal of Military Psychology*.

Schreiber, B. T. & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Summary report.* (AFRL-HE-AZ-TR-2006-0015-Vol I). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schreiber, B. T., Gehr, S. E., & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Real-time and blind expert subjective assessments of learning.* (AFRL-HE-AZ-TR-2006-0015-Vol III). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schreiber, B. T., Rowe, L. J., & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Participant utility and effectiveness opinions and ratings.* (AFRL-HE-AZ-TR-2006-0015-Vol IV). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schreiber, B. T., Watz, E. A., & Bennett, W. Jr. (2003). Objective human performance measurement in a distributed environment: Tomorrow's needs. In *2003 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

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 Behavioral Representations in Modeling and Simulation (BRIMS) Conference*. Scottsdale, AZ.

Symons, S., France, M., Bell, J., & Bennett, W., Jr. (2006). Linking knowledge and skills to Mission Essential Competency-based syllabus development for Distributed Mission Operations. (AFRL-HE-AZ-TR-2006-0041, ADA453737). Mesa AZ: Air Force Research Laboratory, Warfighter Readiness Research Division.

Watz, E. A, Keck, L., & Schreiber, B. T. (2004). Using PETS software to capture complex objective measurement data from Distributed Mission Operations (DMO) environments. In *Proceedings of the 2004 Spring Simulation Interoperability Workshop*. Arlington, VA. Paper available as 04S-SIW-143.

Watz, E. A., Schreiber, B. T., Keck, L., McCall, J. M., & Bennett, W. Jr. (2003). Performance measurement challenges in Distributed Mission Operations environments. [Electronic Version] *Simulation Technology Newsletter.* 03F-SIW-022. March 16, 2004. retrieved 7/26/04 from http://www.sisostds.org/webletter/siso/Iss_102/index.htm 2003 Fall SIWzie Award winner - Fall Simulation Interoperability Workshop.

ACRONYMS

AA   Air-to-Air
AFRL   Air Force Research Laboratory
AIM   Air Intercept Missile
AMRAAM   Advanced Medium Range Air-to-Air Missile
ANOVA   Analysis of Variance
ATES   Automated Threat Engagement System
AWACS   Airborne Warning and Control System
CAF   Clear Avenue of Fire
DCA   Defensive Counter Air
DIS   Distributed Interactive Simulation
DLZ   Dynamic Launch Zone
DMO   Distributed Mission Operations
FLO   First Launch Opportunity
HLA   High Level Architecture
IEEE   Institute of Electrical and Electronics Engineers
IOS   Instructor Operator Station
MAR   Minimum Abort Range
MEC   Mission Essential Competencies
MOR   Minimum Out Range
MSL   Mean Sea Level
MTC   Mission Training Center
N-Pole   Notch-Pole
OCA   Offensive Counter Air
PETS   Performance Effectiveness/Evaluation Tracking System
Pk   Probability of kill
SADL   Situation Awareness DataLink
SCU   Software Capabilities Upgrade
SME   Subject Matter Expert
TSPI   Time, Space, and Positional Information
USAF   United States Air Force
WEZ   Weapons Engagement Zone
WIC   Weapons Instructor Course

APPENDIX: EXPLANATION OF ESCAPE-G

When measuring the performance of the air combat pilot, outcome measures such as kill ratios and missile hit ratios are obviously a first concern. However, as these outcome measures only provide the end result, they are infrequent and do little to reveal how well a pilot is performing during an engagement. Sensitive and more readily available performance measures would allow theoretical research examination of possible discriminating behavioral events or evaluation of Mission Essential Competency (MEC) skills (Colegrove & Alliger, 2002) such as controls intercept geometry or weapons engagement zone management.

Much of the fighter pilot's success or failure rests upon the ability to put his or her aircraft into a region/geometry of opportunity to advantageously employ specific ordnance against a threat while simultaneously trying to deny the threat that same opportunity. That is, a primary goal during air combat is to keep the fighter pilot in an offensive position that greatly increases the probability of weapon intercept with the threat, while simultaneously trying to keep the threat and its weapon's probability of intercept to the friendly quite low.

Theoretical Instantaneous Probability of Weapons Intercept (TIPWI) as an assessment of Weapons Engagement Zone (WEZ) penetration. The WEZ is a relatively simple way for a pilot to think about how far a weapon can travel to a target. It can assist in the cognitive assessment of whether or not a targeted threat aircraft is within a vulnerability zone in order for the pilot to engage. The WEZ is purely a theoretical construct based on the capabilities of the weapon and interaircraft geometries. Since the WEZ, without specific targets, is a theoretical construct, a method to estimate the degree of WEZ penetration by an adversary is desirable. That is, a calculation for the TIPWI was sought (Figure 1). The WEZ introduces an idea of the weapons intercept, while the TIPWI brings an exact and dynamic calculation of the weapons interception probability.

A goal of the fighter pilot in the air-to-air arena is to maintain a TIPWI advantage, which directly contributes to the theoretical probability of kill (Pk). TIPWI is not perfectly synonymous with Pk, but it is a very close approximation of a theoretical Pk and can generally be thought of as such. Most precisely defined, TIPWI is an *ongoing theoretical probability of the weapon intercepting its target if the pilot were to select and launch that specific weapon at that specific threat at that precise moment*. Through the moment of weapon launch, TIPWI is a very precise estimate of Pk. After missile launch, however, TIPWI as an estimate of Pk is no longer appropriate for the missile in flight, only for the missiles remaining on the plane. TIPWI assumes at launch that the missile will then fly-out without failure according to its performance envelope, and it also assumes the pilot will not "trash the missile." Possible affecting factors include, during a radar missile fly-out, a failed control surface, a guidance system failure or other pilot behaviors may lead to a reduced theoretical Pk. In these examples, the TIPWI may have been quite high up until and when the shot was taken, but the Pk could be dramatically reduced during missile fly-out.
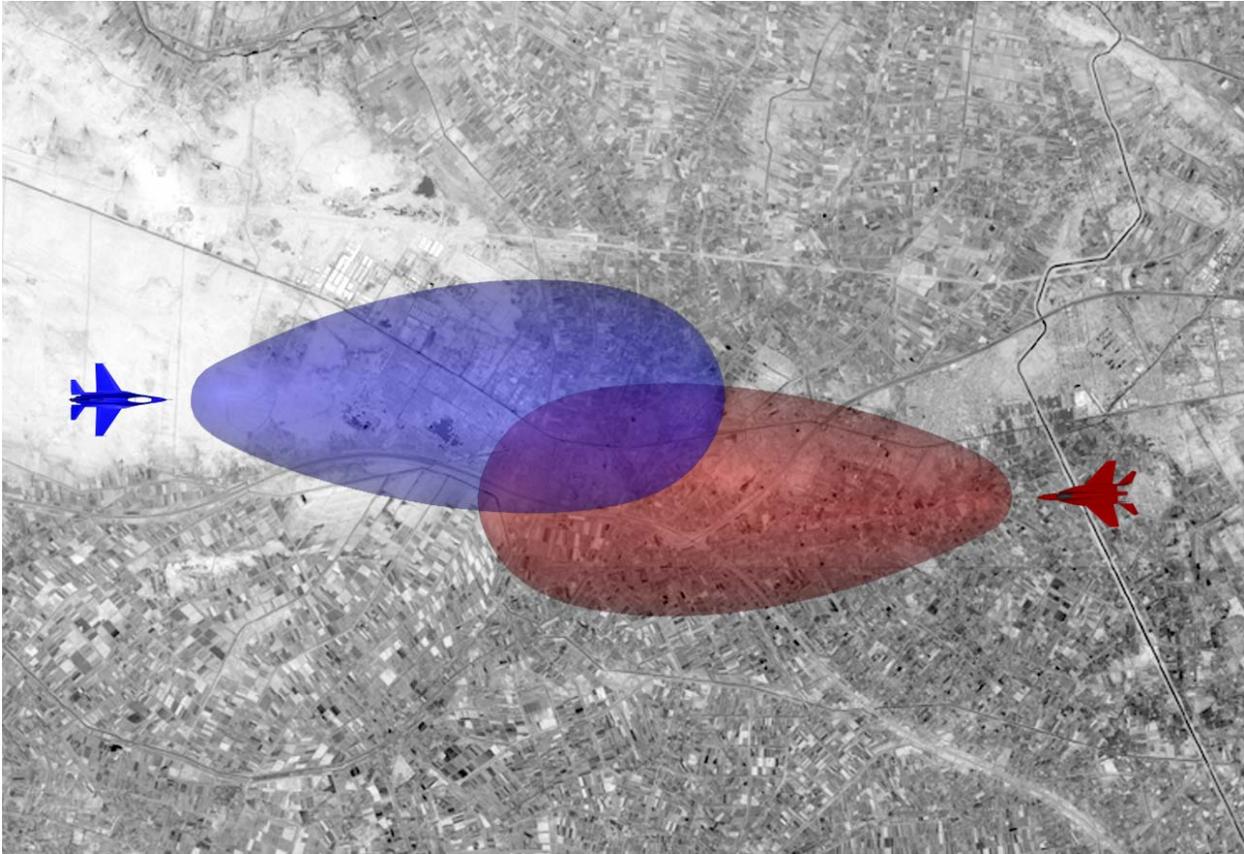
Figure 1 Theoretical WEZ depictions between two aircraft

Due to the fact that no weapon launch is necessary for deriving TIPWI, it can be estimated continuously. TIPWI to a given threat changes constantly and oftentimes quite rapidly, depending upon the moment-to-moment change in inter-aircraft geometry and weapons remaining onboard the firing aircraft. Before coming within critical ranges to a threat, a pilot is trying to use inter-aircraft geometries to his or her advantage in anticipation of raising TIPWI to employ a specific weapon.

Pilot's display of TIPWI estimate. Once within critical ranges, an estimate measure of a pilot's offensive TIPWI opportunity can be presented on the Heads-Up Display (HUD) or avionics display. When the pilot "locks onto" a threat, he or she receives a WEZ display (see Figure 2). The caret position within this WEZ provides the pilot with an indication of the current offensive estimate value of TIPWI to that threat. In this manner, the caret provides the pilot with an ongoing estimate of TIPWI before a shot. The pilot relies heavily upon this TIPWI estimate to decide when and whether or not to shoot. It is important to emphasize that this caret estimate of TIPWI is only displayed to the pilot if the pilot chooses to lock onto a threat. Also, the caret/WEZ is only displayed for that one threat for the one missile chosen. Therefore, before targeting any one threat, the pilot must perform several tasks. Tasks include using the radar, interpreting the radar, evaluating missile capabilities, and positioning the aircraft based upon cognitive estimates of the inter-aircraft geometries and TIPWI values now and in the future.

Presenting a display showing the ongoing TIPWI values to all threats at all times for all weapons would probably greatly aid the pilot and reduce cognitive processing, but such a display is not available in the cockpit. TIPWI is an ongoing composite indication of many of the most important air-to-air combat factors directly related to Pk. It is therefore desirable to maintain high levels of offensive TIPWI over substantial segments of the engagement. TIPWI serves as an ideal candidate for a sensitive real-time performance measure indicative of the air combat pilot's expertise.

Instantaneous geometry between two aircraft and the type of weapons employed are the two critical components in determining TIPWI at any given moment. The instantaneous geometry of the two aircraft defines the current situation and includes numerous factors such as velocity vectors, X, Y, Z positions, relative heading and altitudes, etc. Given an instantaneous inter-aircraft geometry, the probability of intercept is then determined by the weapons onboard the aircraft. With every different type of air-to-air weapon available, any single instantaneous inter-aircraft geometry reveals an applicable portion of each different weapon's performance envelope resulting in different TIPWI values. As an obvious example, the gun does not have the same capabilities as a medium range radar missile. When determining the probability of weapon intercept, it is therefore necessary to continuously calculate what the probability of intercept theoretically would be for each different weapon currently onboard the aircraft to each threat of potential interest.

(R1 or R-maximum)

Caret

(R2)

**No-Escape Zone bounded by R2, R3**

(R3)

(R4 or R-minimum)
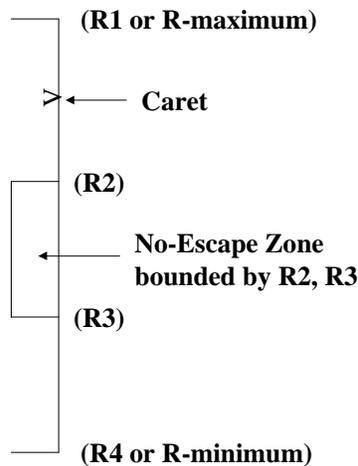
Figure 2 HUD of the AIM-9 Weapons Engagement Zone (WEZ). Caret position and overall display changes according to changes in instantaneous inter-aircraft geometry

Estimating TIPWI using AAMI. The All-Aspect Maneuvering Index (AAMI) is a composite estimate measure of TIPWI developed by Vreuls Research Corporation (1987) and uses the following formula:

$$AAMI = F(ATA) * WRM$$

Where ATA is the Antenna Train Angle and WRM is the Weapons Range Model (Utilizes AOT and Range; refer to Figure 3).
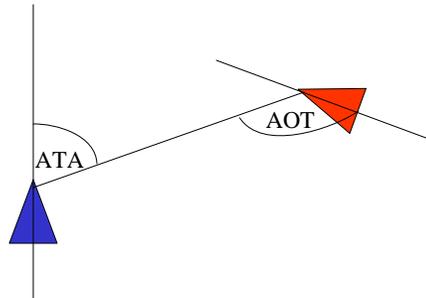
Figure 3 Antenna Train Angle (ATA) and Angle off Tail (AOT) Geometries.

The F(ATA) term is defined as follows: If the ATA is greater than 90 degrees, then F(ATA) equals zero and the AAMI will equal zero. For all other ATA values, F(ATA) = 100[(90 – ATA)/90]. The WRM described next does not use closure velocity (a critical factor for estimating TIPWI), making the F(ATA) term necessary to provide a very rough scaling term to adjust for this factor. The WRM term, or weapons range model term, is derived by using the AOT and range to the threat. For every degree of AOT, four ranges are provided in look-up tables for that particular weapon. The tables provide the maximum range (R1), the minimum range (R4), and the no-escape ranges (bounded by ranges R2 and R3) for a given weapon. If, for a given AOT, the look-up tables provide a value that either exceeds R1 or is under R4, the WRM term will be zero and the AAMI value will be zero, while any value falling between R2 and R3 will yield a WRM term equal to 100. Any AOT yielding a range value from the look-up tables falling between the bounds of R1 and R2 or between R3 and R4 is converted into a value between 0 and 100 by using linear interpolation (see Figure 4). That is, if the R-value derived table look-up is half the distance between R1 and R2, then the WRM term would be .50.

The resulting AAMI term is always a value from 0 to 100, representing an estimate of TIPWI. A zero represents a 0% theoretical chance of weapon intercept, while 100 represents a 100% theoretical chance of weapon intercept. All calculations from the friendly to the threat would be considered offensive estimates of TIPWI, while all the same calculations of the threat to the friendly would be considered defensive estimates of TIPWI (and would result in values ranging from 0 to –100).
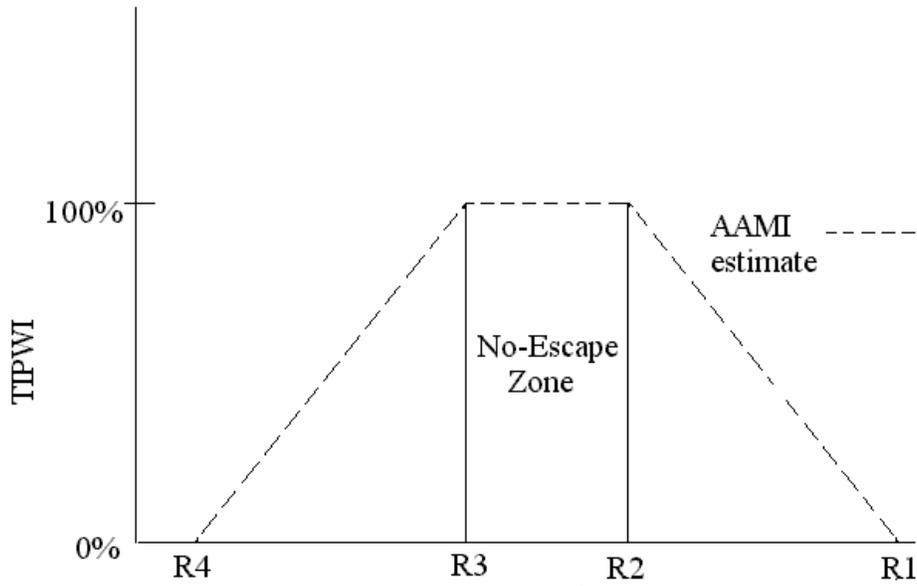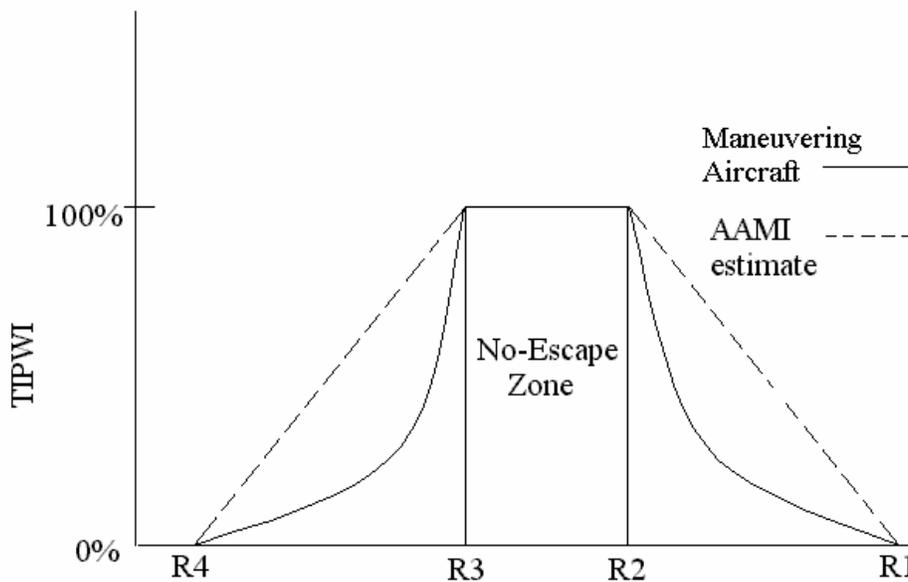
Figure 4 AAMI Estimate of TIPWI as a Function of R-value

The idea of estimating TIPWI by using the AAMI was quite innovative and insightful. A later revision to the AAMI model included additional WRM look-up tables to include altitudes and closure velocities—two critical factors left out in the original model that greatly improves using AAMI as an estimate of TIPWI. Current developmental research at the Air Force Research Laboratory in Mesa, AZ (AFRL/HEA) seeks to provide even a better estimate of TIPWI by adding to and refining the original ideas from the AAMI.

AAMI is based on the weapons of the era in which it was developed—short-range radar missiles, heat-seekers, and guns. Estimates of TIPWI need to include weapons models for today's medium range radar missiles. Most air combat engagements today are won or lost in the radar missile environment that is beyond visual range. Including weapon models for each type of weapon available onboard each aircraft is critical to the success of TIPWI as a performance metric; once provided with a given inter-aircraft geometry, which weapon available is the determining factor for TIPWI. To refer back to our extreme example, if the adversary only has guns remaining while the friendly has medium-range radar and heat-seeking missiles available, the offensive TIPWI values will be dramatically higher for the friendly at almost all inter-aircraft geometries. As a more subtle yet equally important difference, given the same instantaneous long-range inter-aircraft geometry, an older version of an adversary's radar missile could have a substantially lower TIPWI than the newest radar missile version and these differences within weapon type could influence tactics and engagement outcome.

Another imprecision in estimating TIPWI, the AAMI assumes a linear interpolation for any R-value falling between R1 and R2 or between R3 and R4 (see Figures 5 and 7). While R-minimum is based upon a minimum missile time of flight, the caret R-value and the other cut-off R-values of R1-R3 must be based upon a maneuvering assumption by the threat (it is not

31

possible to derive caret or R1-R3 values without a threat assumption.). If the threat were completely non-maneuvering, R-maximum would increase and the entire range between R-minimum and R-maximum would yield a TIPWI of 100% (as the threat is assumed to maneuver less, R2 and R3 would spread outward towards R1 and R4, respectively). A non-maneuvering threat is of course unrealistic; the implied assumption to the linear interpolation procedure used in the AAMI is a very slight, relatively ineffective defensive maneuver by the threat. To calculate and display the WEZ/caret in most modern fighters, the assumption is a high maneuvering threat (e.g., a 6 g-force drag maneuver), a curvilinear function involving dynamic pressure calculations. The TIPWI curves for this assumption would look similar to those depicted in Figure 5. Since the pilot's WEZ display in a real fighter depicts values representing a maneuvering aircraft, that same assumption will be used in our calculations of TIPWI. By doing this, our TIPWI calculations will be based upon the same assumptions used in the real jet and will reflect the same data the pilot receives on the HUD.

Figure 5 Linear Interpolation (AAMI) and Dynamic Pressure Curves (Maneuvering Aircraft) Used in Estimating TIPWI as a Function of R-value. Linear interpolation would always overestimate TIPWI between R1/R2 and between R3/R4 if the assumption were a maneuvering threat.

A third limitation for estimating TIPWI with AAMI calculations is that the AAMI calculations have built-in estimates that are too high. The model allows for estimation of TIPWI at relatively large angles of AOT and ATA. At ATAs approaching 90 degrees, the pilot realistically cannot employ ordnance. To realistically employ a missile, the threat must be within the radar search range, the azimuth and elevation angles of which are limited. Therefore, in our refined calculations for TIPWI, azimuth and elevation angles exceeding the radar limits of that particular aircraft will automatically receive a TIPWI value of zero (for the weapon would have no specific threat to look at). Also, beyond certain g-loading limits of the aircraft, a missile cannot realistically (i.e., should not) be launched because of potential damage to the wing. So, even if

an estimate of TIPWI is slightly positive given the instantaneous inter-aircraft geometry, an estimate of TIPWI should be zero if a given g-loading is also exceeded at that moment.

It is unknown whether the AAMI estimates of TIPWI omitted important factors included in the weapons model or even if threat weapons models were used to calculate an estimate for the defensive TIPWI scores. The weapon models used today at AFRL/HEA are available for both friendly and adversary missiles and include numerous variables regarding instantaneous inter-aircraft geometry, such as potential launch airspeeds that are subsonic, supersonic, or involving shock waves in the transonic region for each type of weapon—a significant factor affecting TIPWI values differentially for various missiles. Though the flight and weapons models in use today at AFRL/HEA are undoubtedly of higher fidelity than those used during the AAMI era, the models could always be improved upon.

Finally, real-time calculations used during the AAMI era were limited to one enemy against one threat, almost certainly due to computer limitations at the time. To improve upon the TIPWI potential application, AFRL/HEA performs all the TIPWI calculations for all entities and all weapons in real-time. This entails a rapid (20 hertz), ongoing measurement of all pairwise inter-aircraft geometries for all friendlies to all threats plus all threats to all friendlies, then accessing all appropriate weapons tables (held in memory), and finally calculating the TIPWI values from these two steps.

Using Escape-G as an estimate of TIPWI. For our estimate of TIPWI, we use a measurement referred to as Escape-G. Escape-G uses the same algorithms the jet uses to display the WEZ, but makes more calculations to determine the precise degree of WEZ penetration by the adversary. That is, the calculations and weapons fly-outs are re-run multiple times per frame taking into account the previously mentioned factors that impact the true value of TIPWI (exact aspects, ranges, altitudes, etc.). Pk for the Escape-G measure therefore very closely follows the theoretical, curvilinear TIPWI Pk estimate lines shown in Figure 5. Escape-G can be thought of as the measure of G-force a pilot must turn while maintaining the same airspeed (to either 0 or 180 aspect, whichever is more appropriate) to defeat the firing aircraft's missile. This is in essence how quickly the pilot must turn at that precise moment if the given weapon were pickled. Fractional readings just above zero (e.g., 0.2) would indicate that the aircraft has just breached the outer boundaries of that weapon's envelope and a shot would be considered a longer range one with a positive, but somewhat lower Pk given the instantaneous inter-aircraft geometries (i.e., an aware and defensive maneuvering threat would likely survive). Readings beginning to approach the limits of human and/or aircraft tolerance (e.g., 6G), on the other hand, would mean reasonably deep penetration into the WEZ with little chance of survival if the weapon were to be launched (i.e, relatively high Pk shot—R2/R3 in Figure 5). For illustrative purposes, we assign Escape-G values exceeding 15G as indicating that the aircraft has flown into the "heart of the envelope" for that weapon and the shot would be considered as a 100% Pk shot and only missile failure or other extreme circumstance would save the targeted aircraft (if the weapon was pickled).

Escape-G can be utilized as a measure of how well a pilot is managing his or her WEZ during the engagement. A pilot who is successful in managing the WEZ will generally have the threat aircraft in positions of high escape-G values. Adversaries in positions with high Escape-G

values indicate a favorable inter-aircraft geometry for the friendly to the threat given the weapons onboard their respective airframes. Similarly, the adversary desires favorable WEZ geometries and attempts to manipulate the inter-aircraft geometries such that the Viper is in states requiring high Escape-G values. Whichever aircraft maintains the geometric advantage and a high Escape-G value will have the advantage over the adversary.

Colegrove, C. M. & Alliger, G. M. (2002, April). *Mission Essential Competencies: Defining combat mission readiness in a novel way*. Paper presented at the NATO RTO Studies, Analysis and Simulation (SAS) Panel Symposium. Brussels, Belgium.

Vreuls Research Corp. (1987). *Air Combat Maneuvering Performance Measurement System for SAAC/ACMI, Volume II, Appendices 5 & 6.* Wright-Patterson AFB, OH: Air Force Systems Command.