Evaluating the Impacts of Mission Training via Distributed Simulation on Live Exercise Performance: Results from the US/UK "Red Skies" Study

Ebb Smith
Bedford Technology Park
Thurleigh, Beds, MK44 2FQ
United Kingdom

MESMITH@mail.dstl.gov.uk

Heather McIntyre
Bedford Technology Park
Thurleigh, Beds, MK44 2FQ
United Kingdom

HMMCINTYRE@mail.dstl.gov.uk

Sara Elizabeth Gehr
6030 S. Kent St.
Mesa, AZ 85212
United States

Liz.gehr@mesa.afmc.af.mil

Margaret Schurig
6030 S. Kent St.
Mesa, AZ 85212
United States

Margaret.Schurig@mesa.afmc.af.mil

Steve Symons
6030 S. Kent St.
Mesa, AZ 85212
United States

Steve.Symons@mesa.afmc.af.mil

Brian Schreiber
6030 S. Kent St.
Mesa, AZ 85212
United States

Brian.Schreiber@mesa.afmc.af.mil

Winston Bennett, Jr.
6030 S. Kent St.
Mesa, AZ 85212
United States

Winston.Bennett@mesa.afmc.af.mil

ABSTRACT

Over the past several years, the UK Defence Science and Technology Laboratory (DSTL) and the US Air Force Research Laboratory (AFRL) have been involved in research to develop training and assessment methods for use in mission training via distributed simulation. As part of our efforts, we have developed competency-based training and rehearsal scenarios and data collection instrumentation for routinely delivering and assessing distributed training events. Our most recent collaborative study, named "Red Skies," involved extending our work to include field assessments of the training benefits derived from involvement in a simulation–based distributed mission training event and subsequent live flying at a Coalition

Red Flag exercise event at Nellis Air Force Base, Nevada in the US. The event was the largest Red Flag event ever hosted, and presented a number challenges for the study and for data collection. This paper will present the methods and results from the distributed simulation preparation, which involved connecting simulators in Bedford UK with those of the Warfighter Readiness Research Division of the AFRL, located in Mesa, Arizona. We will also present results from the subsequent live fly and evaluation in the Coalition Red Flag exercise. These results will include demonstrating our first and successful attempt to follow coalition pilots from a DMO training event to the Red Flag exercise and to collect data while the live fly event was underway. Outcome data will be presented and discussed. We will close with a discussion of the implications of our work to date as a way forward for future training events and how the methods and process developed for Red Skies can allow us to quantify and demonstrate training benefits from distributed simulation training for live operations.

1.0   OVERVIEW

Trial Red Skies was a synthetic collective training exercise, undertaken as part of the Coalition Mission Training Research programme (CMTR). CMTR is a long-term UK/US/CA collaborative project which has been influential in the development of both the UK Mission Training through Distributed Simulation (MTDS) and the US and CA Distributed Mission Operations (DMO) initiatives. This collaborative activity has been very fruitful, offering unique opportunities for cross-fertilisation of ideas and methodologies appropriate to coalition distributed mission training. It is conducted under the auspices of two groups of the Technical Co-operation Panel (TTCP); the Human Resources and Performance Group, Technical Panel 2 (HUM TP-2, Training Technology) and the Aerospace Systems Group, Technical Panel 1 (AER TP-1 Aerospace Operational Analysis and Simulation) via a long-term Project Agreement (PA).

Red Skies was the third CMTR trial to be undertaken since November 2001. In line with the PA, the trial was designed to address the following broader CMTR research objectives

- Develop systems to mitigate the effects of interacting real-time simulators caused by extreme long distance links.

- Design processes for creating scenarios to fulfill specified training objectives and develop metrics for measuring impacts on learning.

- Develop systems for distributed mission planning, briefing, replay and debriefing.

- Create assessment tools for evaluating the effectiveness of distributed simulation for enhancing warfighter skills in conducting coalition force operations.

2.0   DESCRIPTION OF RED SKIES

Red Skies took place during the week 28th February to 4th March 2005. The trial was designed to support training for aircrew prior to their participation in the March 05 Exercise Red Flag (Joint Red Flag 5-3.2). The trial involved operational UK Tornado GR4 aircrew (from 13 Squadron RAF Marham), E-3D aircrew (from Air C2 Operational Evaluation Unit, RAF Waddington) and USAF F-16 aircrew (from 113st Fighter Squadron, Terra Haute, Indiana).

The synthetic environment for this exercise included virtual and computer generated Red and Blue forces within a networked, operationally realistic and highly dynamic scenario comprising IADS, EW and a real-world terrain data base of the Nellis Air Force Base (AFB) Range Complex, Nevada. A secure network link

between the UK Aircrew Training Research test-bed at QinetiQ, Bedford and the US Air Force Research Laboratory (AFRL) at Mesa, Arizona enabled the UK and US aircrew to plan, brief, fly and debrief during a week long simulated Red Flag exercise. Staff from the Air Warfare Centre Tactical Team provided White Force support to optimise the operational value and provide the Exercise Management function in the UK.

Air Interdiction (AI), Close Air Support (CAS) and Time Sensitive Targeting (TST) missions were flown each day and the White Force included two front-line Forward Air Controllers (FAC) from Delhi Barracks, Tidworth. In real-world operations they provide the FAC element of the Tactical Air Control-Party (TAC-P).

The CMTR team also attended the live Joint Red Flag 5-3-.2 at Nellis AFB, Nevada, to complete the data gathering needed to undertake a transfer of training study. As in Red Skies, the crews flew both AI and CAS/TST missions. The major differences were the number of participants (over eighty aircraft) and both day and night missions were flown.

Post trial analysis indicated that the Red Skies transfer of training trial to the live Red Flag was a success with positive feedback from all participants.

3.0   TRIAL AIM AND EXPERIMENTAL DESIGN

As described, Red Skies was a CMTR event that was planned around Joint Red Flag 5-3.2. The goal of this event was to investigate the training value of simulation based exercise preparation as part of the buildup to Red Flag. Since this Red Flag exercise was the largest ever involving many coalition partners as well as U.S. forces, it was deemed critical to the design of the virtual Red Skies CMTR event to fly missions as similar to those anticipated during Red Flag as possible. This included mission planning via video teleconference, a coalition package emulation likely to occur during Red Flag, an Air Tasking Order (ATO) including all players, Red Flag Special Instructions (SPINS) and training rule adherence, and flying the missions over the Nellis range complex and target areas. To accomplish this, AFRL/Mesa established a relationship with the Red Flag personnel in charge of these areas and was able to receive the latest versions of these documents, as well as the data from the Red Flag planning conferences to include ATO targets and taskings. Additionally, AFRL/Mesa was able to coordinate the ATO during Red Flag to ensure that the US and UK aircrews that participated in Red Skies were able to fly together in Red Flag to the max extent possible to include the days that they were going to be mission commander. This gave the labs the greatest amount of leverage to design and construct a study to capitalize on these control measures. Performance assessment and training transfer were also key objectives in this exercise so it was critical that a common set of data research protocol was adhered to in addition to a certain amount of rigidity in the design study. In addition to the data collected before, during and after the Red Skies event, the F-16 pilots that flew at AFRL/Mesa also participated in benchmark missions that allow them to be compared to the cohort data set that has been established through 3 plus years of training research. Table 1 shows the template used to maximize performance documentation and training transfer in the design study:

Table 1: Table of objectives

Players	Training AG benchmarks and AG type syllabus	Metrics	Control	Metrics	Possible enroute benchmarking	POST EX
F16:12 pilots/8 jets	5 pilots Mesa	Demographics, MEC surveys; Pathfinder Top 3/Bot 3; PETS data	7 Home station ops; HS sortie tracking	MEC surveys; daily top 3/Bot3	All to Mesa before Red Flag; MEC surveys; Pathfinder Top 3/Bot 3; PETS data	MEC surveys; PETS range data
GR4: 24 pilots/8 jets	8 pilots Bedford	Demographics; MEC surveys; Pathfinder; Top 3/Bot 3	16 Home station ops; HS sortie tracking	UK MEC surveys; daily top 3/Bot 3	N/A	MEC surveys; range data
UK AWACS:	3 Mesa	Demographics; MEC surveys; Pathfinder; Top 3/Bot 3; Recorded data	10 Home station ops; HS sortie tracking	UK MEC surveys; top 3/Bot 3	MEC surveys; Pathfinder Top 3/Bot 3; Recorded data	MEC surveys; range data

4.0   METHODS OF ASSESSMENT AT RED SKIES AND RED FLAG

One of the main goals for the assessment of Red Skies was to continue to use an established set of assessment tools and protocol [1] that was developed for Exercise First WAVE. The opportunity existed to take the lessons learned from and refinements made after EFW and continue the evolution of the assessment tools and protocol. The more streamlined approach worked very well and we were successful in gathering assessment data in pre-Red Skies, Post-Red Skies, pre-Red Flag and Post-Red Flag environments. One important lesson from our EFW experience was the need to reduce and fine tune the survey instruments and data collection activities to make the process less intrusive without losing the linkage of the methods and questions to the overall objectives of each training event and activity. Our EFW experience underscored the absolute importance of gathering systematic data from individuals and teams and merging those data with objective data gathered from the simulation environment. While some of our EFW partners expressed concern for the scientific approach we collectively decided to take and were concerned about the comprehensive nature of the process and protocol, the convergence of findings on the overall success of EFW, the adoption of the standard protocol and instruments, and the collective development of the common instruments and language, is unquestionable. For Red Skies, we made substantive refinements to both the content and the coverage of the instruments to ensure their breadth of coverage and their uniform applicability for future events. Table 2 illustrates the surveys and sources of data collected during Red Skies and Table 3 shows the same for Red Flag.

TABLE 2: Surveys and instruments used for Red Skies data collection.

Instrument: Survey	Primary Questions Addressed	Rationale	Target Respondent Groups	Primary Analysis Goals
Pre Red Skies exercise data collection				
• Demographic	What training environments and hours have you had prior to today?	• Background knowledge about participants	• A/G pilots • AWACS	• Sort participants by experiences
• Experiences and Learning Environments (Frequencies)	How often do pilots engage in each experience within each training method (e.g., live fly, combat, simulator, other training device)?	• Determine current experience load for each training/learning environment	• A/G pilots • AWACS	• Baseline gap analysis for each nation and for MTDS overall • Identifies most common experiences • Frequency of experience in current weapon system
• Experiences and Learning Environments (Ratings)	To what extent can a pilot engage in each experience within each training method (e.g., live fly, combat, simulator, other training device)?	• Determine ideal training environment for different experiences	• A/G pilots • AWACS	• For each experience, environment(s) providing experience
• Mission Specific Expectations	A-priori, what are the expectations each operator has regarding the effectiveness of distributed training for each experience?	• Determine what the participant's expectations are before exposure to First WAVE	• A/G pilots • AWACS	• Determine common expectations
Within Red Skies exercise data collection				
• Mission Process Assessment	What is the assessment of the exercise performance each day throughout Red Skies, (linked to supporting competencies)?	• Have a SME evaluate pilot performance each day to track improvement	• WFLO/SME that has background expertise in the role of the operator	• Track training through the event
• Engineering and Behavioral Observer Gradesheet	What were the specific problems in each of the following areas?	• To ensure a record of problems is kept to decrease problems in future events	• Engineers • Observers • Researchers • WFLO	• Provide a record of problems and solutions
• Daily summary of "top 3" and "bottom 3"	What were the best 3 and worst 3 events that occurred today?	• A quick look at the good and the bad from each day of the exercise	• All participants	• Track common problems
• WFLO observational survey	From the unique WFLO perspective, how did the various phases of the mission go?	• Provide knowledge for future WFLOs	• WFLO	• Find common problems and solutions

Instrument: Survey	Primary Questions Addressed	Rationale	Target Respondent Groups	Primary Analysis Goals
Post Red Skies exercise data collection				
• Mission Specific Experiences	What is the perception of training effectiveness for each experience resulting from their participation in Red Skies?	• Determine if the pilots could have certain experiences in the coalition environment,	• A/G pilots • AWACS • Mission Commander	• Is it possible to train different experiences in the coalition environment?
• Attitudes towards Red Skies and Mission Training through Distributed Simulation	What opinions did the participants have about the critical constructs of the exercise?	• Determine if the pilots thought that the training had value	• A/G pilots • AWACS	• Determine overall positive acceptance of the exercise
• WFLO Interview	How well did Red Skies run?	• Give the WFLOs a final chance to give their opinions about the exercise	• WFLO	• Gather opinion data of WFLOs to improve future coalition events

Table 3: Surveys and instruments used in Red Flag data collection

Instrument: Survey	Primary Questions Addressed	Rationale	Target Respondent Groups	Primary Analysis Goals
Within Red Flag exercise data collection				
• Mission Process Assessment	What is the assessment of the exercise performance each day throughout Red Skies, (linked to supporting competencies)?	• Have a SME evaluate pilot performance each day to track improvement	• WFLO/SME that has background expertise in the role of the operator	• Track training through the event
• Engineering and Behavioral Observer Gradesheet	What were the specific problems in each of the following areas?	• To ensure a record of problems is kept to decrease problems in future events	• Engineers • Observers • Researchers • WFLO	• Provide a record of problems and solutions
• Daily summary of "top 3" and "bottom 3"	What were the best 3 and worst 3 events that occurred today?	• A quick look at the good and the bad from each day of the exercise	• All participants	• Track common problems
Post Red Flag exercise data collection				
• Mission Specific Experiences	What is the perception of training effectiveness for each experience resulting from their participation in Red Skies?	• Determine if the pilots could have certain experiences in the coalition environment,	• A/G pilots • AWACS • Mission Commander	• Is it possible to train different experiences in the coalition environment?
• Attitudes towards Red Skies and Mission Training through Distributed Simulation	What opinions did the participants have about the critical constructs of the exercise?	• Determine if the pilots thought that the training had value	• A/G pilots • AWACS	• Determine overall positive acceptance of the exercise

For Red Skies and Red Flag we gathered both subjective ratings and objective measures of performance. In addition, it was critical for the focus of the Red Skies training to parallel the training objectives of the Joint Red Flag as much as possible. Given that one of the goals of Red Skies was to evaluate the potential for tracking simulation training transfer to the operational event, we needed to ensure objectives matching

between the two events. Table 4 shows the match of objectives and scenario characteristics from the Red Skies activities and those in Red Flag. It was also important to make sure that the Red Flag support cadre was able to schedule (task) the USAF and UK crews to fly together several times over the two weeks of the Red Flag event so that we would have opportunities to directly observe live fly performance that was matched to similar activities in the Red Skies event. This manipulation of the actual air tasking orders to support a training research program is a significant advance in bringing the training and operational exercise communities together to help demonstrate the payoffs of one for the other. Our experience in Red Skies and Red Flag was a tremendous step forward in bringing training and operations much closer to one another. We feel that the results from Red Skies and Red Flag demonstrate the potential for advanced simulation as a preparation for exercises (as surrogates for combat mission performance) and eventually for actual combat preparation.

TABLE 4: Experience and objectives match

EXPERIENCES	RED SKIES	RED FLAG
USAF / UK Coord / Similar Package Comp	X	X
ATO Drop / Mission Planning	X	X
Complex Ground Operations	•	X
Air Refueling Operations	•	X
Nellis Data Base / Nellis Range Complex	X	X
Nellis Bullseye (Cedar Peak)	X	X
Nellis Target Arrays (60 and 70 Series)	X	X
Nellis Range Restrictions	X	X
Nellis Air-to-Air Threat Array	X	X
Nellis Surface-to-Air Threat Array	X	X
RED FLAG Training Rules	X	X
RED FLAG SPINS	X	X
RED FLAG Dynamic Targeting Procedures	X	X
RED FLAG Recovery Procedures	•	X

4   RESULTS

4.1   Overall Results

This section will present a brief overview of the surveys and results from Red Skies and the associated Red Flag. The aircrew that participated in Red Skies and Red Flag ranged in qualification from wingman to mission commander. Before Red Skies, they completed a survey that asked them to rate how effective they expected Red Skies to be at training different experiences derived from Mission Essential Competencies (MECs) [2]. The rating scale ranged from 0 (not at all effective) to 4 (very effective) with an option for Does Not Apply if they felt that that experience could not be replicated in Red Skies. At the end of the week, they filled out a similar form asking how effective Red Skies was at training the different experiences. At the end

of Red Flag, they again filled out a survey asking how effective the overall Red Flag spin up was at training the different experiences.

Subject Matter Experts (SMEs) rated the aircrew's performance each day during Red Skies and Red Flag. The rating scale ranged from 1 (performance indicates a lack of ability or knowledge) to 5 (performance reflects an unusually high degree of ability) with an option to select N/A if they had no opportunity to observe the pilots in this area.

At the end of Red Skies, all aircrew were also asked to complete a survey about their attitudes towards Red Skies and mission training through distributed simulation. They were given a series of statements and asked to rate their agreement on a 4 point scale from strongly disagree to strongly agree. They were given a similar survey at the end of Red Flag that asked them about their attitudes toward the overall Red Flag spin up. The average results of these surveys across all aircrew (and all days for the performance survey) are summarized in Table 5.

	Red Skies	Red Flag
expectations	2.15	
experiences	1.73	1.99
performance	2.95	3.46
attitudes	2.76	2.68

TABLE 5: Red Skies and Red Flag survey results

Finally, at the end of each day of Red Skies and Red Flag, all aircrew were given an open ended survey that asked them to list the top 3 and bottom 3 events from the day. This gave them an opportunity to comment on any phase of the mission, from brief, through execution, to debrief. During Red Skies the comments from the aircrew reflected high potential for such exercises. One comment was "The mission flown provided a challenging look at a Red Flag / opposed sat profile. The opportunity to train with threat emmiters was one of the high points." On Thursday, one aircrew went so far as to say "Great experiences to take w/me to Red Flag" showing high pilot acceptance of programs like Red Skies.

We collected objective data both in the simulators and with live-fly using software tools and objective measures. Initial technical challenges prohibit us from reporting data in this manuscript, but it does appear that the capability to do this routinely will exist in the relatively near future, thereby enabling objective transfer of training research on the exact same selected mission outcome and process (i.e., skill) measures of performance.

4.2   Post Exercise interviews

Post Red Flag, interviews were also conducted with Red Skies participants to capture feedback on training value and to confirm transfer of training. Summaries of the interviews undertaken with UK participants are given below.

4.2.1   UK AWACS Weapons Controller team

This was the first time UK AWACS aircrew had been involved in a CMTR trial and they were very enthusiastic about the training value of Red Skies. The AWACS weapons team confirmed that debriefs tended to cover operational issues rather than 'simisms'. This is a very positive sign. Full mission replay with voice

also provided incontrovertible evidence of what was actually said during a sortie. The debriefs were deemed to be very representative of Red Flag; tactical, effective and honest.

The team were extremely enthusiastic about the Red Skies experience and the overall MTDS/DMO concept. They felt it offered superb training and any criticisms are really minor ones. As with participants in previous trials, the first reaction (pre-trial) was one of cynicism, however once the trial started they quickly appreciated the real training benefits it offered. All the UK weapons team agreed that Red Skies was the best synthetic exercise they had ever been involved with and that they had benefited from the experience. They concluded that Red Skies was very good preparation for the Joint Red Flag

4.2.2   UK Forward Air Controller

The UK FAC confirmed that Red Skies provided useful training in terms of interaction with US aircraft. It also provided useful interactions with others elements (role-played by the WF) that FACs would not usually have any exposure to. It offered FAC training opportunities for ML CAS prior to going on major exercise

4.2.3   UK Air to Ground Crews

The crews were interviewed shortly after their return from Red Flag. The crews confirmed that Red Skies had helped them prepare for Red Flag, in that it provided a great degree of familiarisation with the complex and restrictive airspace in which Red Flag is conducted. The crews also became very familiar with the range areas, indeed one of the pilots commented that on ingress during an early sortie he remarked to his navigator that he knew exactly where he was. It also gave the participants a greater insight into roles of the various US players such as the SEAD assets, and the more inexperienced crews gained a lot of training benefit from these interactions. The weeks work-up also gave the crews valuable COMAO training which was normally unavailable during normal squadron training and provided valuable pre-deployment Mission Commander training.

5   IMPLICATIONS FOR FUTURE TRAINING EVENTS

There are a number of overall implications of Red Skies-type events for future linkages amongst training and operations. First is the further refinement and extension of the comprehensive training evaluation and data collection protocol developed in EFW and then extended in this event. This common approach, which was developed through a seven-nation collaboration was used successfully in both EFW and Red Skies and will be used again in future events as well. Third, and on the USAF side of Red Skies, the event provided considerable opportunity for us to explore the integration of our objective performance measurement system in a live fly/range environment. In Red Skies, we were able to prove the concept of embedding training performance assessment in an operational context. While this embedding was not without problems, as were discussed earlier in the paper, it was clearly an initial success that bodes well for future, larger scale integration. Finally, the USAF Air Combat Command has taken the data collection protocol and instruments developed for EFW and Red Skies and is using them for evaluation data collection of large scale Distributed Mission Operations (DMO) training events know as Virtual Flags.

As anther example, other forces, in collaboration with USAF and UK researchers and operations personnel, are embarking on a "Maple Skies" event in the spring of 2006 where the Red Skies model will be applied to preparations for and participation in a Maple Flag event. We are implementing a similar protocol, data collection, and evaluation approach using the assessment data base and instruments we used in EFW and in Red Skies/Red Flag. With Maple Skies, we are seeing the continued institutionalization of a more unified

protocol and set of objectives applicable to both simulation based preparation training, and assessment of training impacts on operational mission performance in live fly exercises. We are also able to identify specific standards for measurement and data control that we have never had in the past. As each nation adds simulation capability, we can expand the protocol and data collection activities to encompass the new capabilities and grow the data bases we are creating as well.

As the USAF and our coalition partners grow their capabilities, we can use the common protocol, metrics and standards to develop common data analysis and reporting schemes that can be used within and across future coalition training and exercise events. With Red Skies/Red Flag, we were able to demonstrate the potential for integrating simulation-based training with live operations and to illustrate how to ensure that common objectives are identified and implemented in both environments. These common objectives are a critical component of any future training transfer assessment activity or study. Finally, the lessons learned and approach developed for this event serve as a benchmark for future training transfer evaluations.

6   REFERENCES

[1]   Gehr, S.E., Schurig, M.A., Jacobs, L., van der Pal, J., Bennett, W., Jr., and Schreiber, B.S. (2005). Assessing the Training Potential of MTDS in Exercise First WAVE. Paper presented at Symposium on the Effectiveness of Modeling and Simulation – from Anecdotal to Substantive Evidence, Warsaw, Poland October 2005.

[2]   Colegrove, C. M., & Alliger, G. M. (2002). Mission essential competencies: Defining combat readiness in a novel way. Paper presented at: *NATO Research & Technology Organization, Studies, Analysis, and Simulation Panel, Conference on Mission Training via Distributed Simulation (SAS-38)*, Brussels, Belgium.