2007

Assessing Large-Scale Training Events: Developing Common Tools and Protocols

Sara Elizabeth Gehr

Margaret Schurig

Leah J. Rowe

Ebb Smith

Winston Bennett Jr.

ASSESSING LARGE-SCALE TRAINING EVENTS: DEVELOPING COMMON TOOLS AND PROTOCOLS

Sara Elizabeth Gehr
The Boeing Company
Mesa, AZ

Margaret Schurig
The Boeing Company
Colorado Spring, CO

Leah J. Rowe
L-3 Communications
Mesa, AZ

Ebb Smith
Defence Science and Technology Laboratory
Bedford, UK

Winston Bennett, Jr.
Air Force Research Laboratory
Mesa, AZ

Numerous previous studies have shown the positive effects that simulators provide in training aircrew for both basic maneuvering and more advanced team combat skills. As technology evolves, we have the ability to connect simulators separated by great geographic distance and often in different countries, thus allowing an even greater number of aircrew to train together as a team. In addition, live-fly training events provide an excellent opportunity to train aircrew. However, these large-scale distributed and live-fly training events must be examined for effectiveness. To accomplish this, training assessors must use a common set of data collection instruments, and follow a common protocol that says when to use them, and who should use each one. This is particularly important in distributed simulation exercises where training participants (and assessors) may be located at a number of geographically dispersed sites. The Air Force Research Laboratory along with its international partners has developed a set of knowledge, skill and experience based surveys that can be used to evaluate large scale distributed training events. In addition, we have developed a protocol that specifically describes when each survey should be given, and to whom. These events pose unique challenges in data collection; therefore, in addition to traditional paper surveys, we have developed a website and "minibrowser" that can run on a stand alone, non-networked computer to collect data from participants. These tools, protocols and data collection methods have already been successfully used at a number of large scale distributed events. This paper will present data and results from US and multinational applications of the methods and tools, discuss feedback and lessons learned. We will discuss and highlight the latest versions of the tools, protocol and methods we have developed, and present the future directions we hope to take with this research

Introduction

Routine assessment of training that occurs in geographically dispersed locations presents several challenges. First of all, as the number of sites increases, the probability that there may not be a person at each site who is knowledgeable about the assessment process increases. The lack of knowledgeable people leads to increased chances that assessment instruments will not be filled out, will be filled out incorrectly, or will be filled out at the wrong time during the training event. It is rare that an aircrew member will ask to ensure that he has completed all necessary surveys. Therefore, it is advisable for someone to be present at each site who is familiar with the assessment goals, objectives, methods, tools and protocols to ensure that the assessment is carried out. Over time as the instruments and protocols become institutionalized the need for proctors who are not local to the site should be substantially reduced.

Another challenge is that each site may have diverse assessment goals. This is expected, especially when the sites are in different countries, each with slightly different objectives for the training event. However, even with different goals, a common set of tools and protocols should be agreed upon during the planning of the exercise. Of course, it would be completely

appropriate for each site to add a limited number of measures in which they are particularly interested, as long as the additions do not interfere with the agreed upon tools and protocols.

Even given these challenges, it is important that all participating sites agree well in advance of the training exercise what tools and protocols will be used. This assessment planning should be tightly integrated with the exercise planning, and will ensure that all sites are using the same tools, protocols, and data collection methods at the same time during the training event.

In this paper we will describe an approach to distributed data collection that we have developed and used during distributed exercises over the past few years. We will outline the development process, the tools and protocol, and describe both implementation and refinement in recent multi-national events.

Description of Tools and Protocols

Our goal in developing tools for assessing distributed training events was to base them on Mission Essential Competencies (MECs). MECs are "Higher-order individual, team, and inter-team competencies that a fully prepared pilot, crew or flight requires for successful mission completion under adverse conditions and in a non-permissive environment" (Colegrove and Alliger, 2002). MECs provide a common framework for developing a common set of assessment tools and methods. MEC definitions come from subject matter experts and data on the combat-relevance of MECs from field data collection activities.

Of particularly note are the MEC based surveys. Currently MECs capture experiences, knowledge, skills and competencies at a mission or role-specific level. In large scale training events (involving a number of different roles and platforms) cross-walking mission or role-specific requirements to identify cross-mission or inter-role competencies is very important. Definition of this level of analysis will potentially identify those roles, and therefore a locus of assessment, that would most benefit from synthetic collective/coalition training, and the optimal mix of manned versus constructive players.

One proposed method for identifying these competencies is this "cross-walking." Cross-walking involves taking the MECs that have been identified for several roles, and looking for commonalties across roles. This cross-walk would enable the identification of things that are oriented towards two or more teams or packages. It is anticipated that once MECs have been identified for several roles, it will be possible to identify commonalties across these roles. The current collaborative MEC-based surveys can be refined to reflect these commonalities. This will facilitate a useful assessment of the large-scale training event in respect of the collective training requirement.

The tools we developed are outlined in Table 1, along with when in the exercise they are typically administered (pre, during or post). This specific example is from Exercise First WAVE (Warfighter Alliance in a Virtual Environment). First WAVE was one of the very first attempts to develop and use common tools. The exercise, which involved six nations linking simulation facilities and devices together, provided an excellent opportunity for this development to occur and for the research team to obtain input from large and diverse research and mission area pool of experts. The tools we have today are a direct result of the multinational collaborative development and implementation that occurred in First WAVE.

Table 1. Common tools developed for distributed training exercises (A/A = Air to Air; A/G = Air to Ground; AWACS = Airborne Warning And Control System; GCI = Ground Control Intercept; WFLO = White Force Liaison Officer; MTDS = Mission Training through Distributed Simulation; SME = Subject Matter Expert)

Instrument: Survey	Primary Questions Addressed	Rationale	Target Respondent Groups	Primary Analysis Goals
Pre exercise data collection				
Demographic	What training environments and hours have you had prior to today?	Background knowledge about participants	• A/A pilots • A/G pilots • AWACS • Red GCI	• Sort participants by experiences
Experiences	How often do pilots engage	Determine current	• A/A pilots	• Baseline gap analysis

Instrument: Survey	Primary Questions Addressed	Rationale	Target Respondent Groups	Primary Analysis Goals
and Learning Environments (Frequencies)	in each experience within each training method (e.g., live fly, combat, simulator, other training device)?	experience load for each training/learning environment	• A/G pilots • AWACS • Red GCI	for each nation and for MTDS overall • Identifies most common experiences • Frequency of experience
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/A pilots • A/G pilots • AWACS • Red GCI	• 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 the training event	• A/A pilots • A/G pilots • AWACS • Red GCI	• Determine common expectations
Within exercise data collection				
Mission Process Assessment	What is the assessment of the exercise performance each day throughout of the training event (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 listed 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	• White Force Liaison Officer (WFLO)	• Find common problems and solutions
Post exercise data collection				
Mission Specific Experiences	What is the perception of training effectiveness for each experience resulting from their participation in the training event?	Determine if the pilots could have certain experiences in the current training environment	• A/A pilots • A/G pilots • AWACS • Red GCI • Mission Commander	• Is it possible to train different experiences in the current training environment?
Attitudes towards the	What opinions did the participants have about the	Determine if the pilots thought that	• A/A pilots • A/G pilots	• Determine overall positive acceptance of

Instrument: Survey	Primary Questions Addressed	Rationale	Target Respondent Groups	Primary Analysis Goals
training event and Mission Training through Distributed Simulation	critical constructs of the exercise?	the training had value	• AWACS • Red GCI	the exercise
WFLO Interview	How well did the exercise run?	Give the WFLOs a final chance to give their opinion about the overall exercise	• WFLO	• Gather opinion data of WFLOs to improve future training events

Data Collection Methods

MECs served as the basis for identifying the knowledge, skills, and experiences of relevance within mission areas, and for the development of questionnaires for assessment. Once the tools and protocols have been established, the methods to be used to collect the data must be determined. The resources (computers, copiers, internet access, etc.) that may or may not be available at each site need to be taken into consideration. It is imperative that the method is flexible and provides back-up plans because each site is likely to have different resources, and may need a different data collection method to fit their needs, even though all participants at all sites will be using the same instruments and protocols. In addition, data collection must be flexible enough so that if a resource that was expected to be ready for use is not available, a back-up method should be ready to go.

Another resource that must be taken into consideration when designing data collection methods for different sites is the availability of on-site, dedicated proctors or data collectors. Researchers should play an active role in the data collection process to ensure data is collected according to the proper protocol. An ideal situation is to have the researchers observe the exercise from start to finish. Managing the expectations of the intended participants is another key to successful data collection. Informing the participants prior to arrival for the exercise about what is expected in terms of data collection, how much time it will take, and for what the data are to be used, should be discussed in detail. In doing this, the researcher can gain the respect of the participants, and ensure that daily surveys are being collected and the research protocol being followed. It is time consuming but enriches the quality of the data received.

One seemingly simple method for managing expectations is to ensure there is time in the daily schedule, at the appropriate times, for data collection. As mentioned earlier, this is one reason why the researchers must be involved from the beginning of the exercise planning. If the assessment time is clearly marked in the schedule for all, including the participants, to see, then they will know what is expected of them from the beginning.

For the data collection tools developed, we have three basics methods of collecting our subjective data, a web-based, a “mini-browser” and paper surveys.

Web-Based

We have developed a data collection website that has web based surveys that can be completed on-line, and printable versions of the surveys for the researchers. This type of data collection tool is good because the participant himself is entering in the data and handwriting issues are not a problem. In addition, it eliminates any problems of different versions of surveys floating around, as everyone from every site is accessing the same centrally-stored online surveys. However, if there is any technical problem with the website or with the computer that the participant is using resulting in data corruption, then the data can be lost since there is no back up of the information entered. Furthermore, one also may potentially face security issues when using the internet. It is important that the proper steps and precautions are taken to ensure the privacy of the participants completing the surveys. The website can also be used for post-exercise data entry by the researchers for sites that could not access the internet at their site.

"Mini-browser"

The mini-browser looks similar to the website, but is a stand alone data collection tool that works on any computer with Internet Explorer installed. It does not require internet access in order to function, as it was developed for use at sites where internet access was not available. It has the same functionality as the website, and can be used for both data collection and post-exercise data entry. Again, while using this method during data collection there is not a sure form of backup. Therefore, there is still some potential for losing data if there is a technical problem. Additionally, this tool must be sent to the site in advance, or brought by the data collector, and installed on any computer that will be used for data collection, which takes more time then using the website. Also, the results are saved on the local hard drive in a folder on the desktop. The on site data collector must copy this folder from every computer that was used during the exercise. These will need to be combined with each other and with data from the website, mini-browsers from other locations, and paper surveys to form a complete database of the survey results.

Paper-based

Often when traveling to remote sites, technology, such as the internet or even computers, may not be as readily available. Therefore, it is often necessary to utilize paper surveys while collecting data. Unlike the other tools described, paper provides a valuable back up if technology fails, so even if planning on using another tool, researchers should always have paper available as a back up. Also, there are no chances of computer technical problems or worries about the correct version of software with paper surveys. Furthermore, paper surveys are familiar to the participants and may be less intimidating. However, paper surveys require much more data entry, including the often onerous task of interpreting handwriting of participants, than the other methods. They also must be stored, and an electronic database to store the data must be created and managed. Therefore, they take far more time to collect and manage than using one of the other two tools previously described.

Previous Implementations

We have implemented the methods and tools discussed above several times for large training events including First WAVE, Red Skies, and Exercise Pacific Link.

Exercise First WAVE

The tools, protocols, and methods described above were first developed for and used during First WAVE in November 2004. First WAVE was a seven nation event that demonstrated the potential of Mission Training through Distributed Simulation (MTDS). An extensive discussion of the data collection and results for this event has been published previously (Gehr et al. 2005). This first use of the tools and methods was also the largest to date, involving eleven sites in six countries. This event demonstrated the necessity of having dedicated assessment personnel at each site, as more accurate data was collected from sites that had personnel whose sole task was to ensure that data was collected at the proper time, by the proper instrument, from the correct personnel.

Red Skies

The second large scale use of these tools and methods was in March 2005 during trial Red Skies. Red Skies was a synthetic distributed coalition US/UK exercise in conjunction with a live-fly Red Flag later that same month (Smith et al. 2005). The tools described above, slightly modified, were used during both the virtual Red Skies event and the live-fly Red Flag. One key to the success of live-fly Red Flag data collection was the presence of dedicated data collection personnel at the Red Flag. This presence provided documentation of unpredicted events, and ensured adherence to the proper data collection protocol. Both the US and UK sent personnel to Red Flag to be on-site for data collection.

Exercise Pacific Link

In November 2005, the US and Australia conducted an unclassified demonstration of low-cost connectivity (Crane et al. 2006). Again, the MEC based surveys were used as a basis for assessment. However, because of the smaller scale nature of this event, and the lack of active duty personnel on the US side, a scaled down version of the tools was used. We omitted the pre-exercise MEC experiences surveys, did not have any SMEs rate the performance of the pilots, and did not do any WFLO surveys. Even with this smaller scale exercise and toolset, the assessment was completed successfully at both ends, as there were dedicated assessment proctors at both sites.

Current Uses and Future Directions

In any large-scale training event it is essential that assessment tools or methods developed support the needs of all the participants; trainees and assessors.

The trainees require feedback on performance (at a team and collective level) to ensure that the right training lessons are learnt. The assessors require a rich and reliable data set to facilitate a detailed investigation into the robustness of the synthetic environment from both training and a technical perspective. This is particularly important when the training event is across distributed sites incorporating simulation systems and components of differing levels of capability and fidelity.

The crux of the problem, from a research perspective, is the need to understand the impact of synthetic collective training on practice/maintenance of mission essential skills and competencies. How tangible is the benefit? Which roles would most benefit from distributed synthetic collective/coalition training? What is the optimal distributed synthetic environment model?

With this in mind, applications and lessons learned from the common assessment tools to date have already resulted in a number of refinements that are currently being used to assess Virtual Flag – a very large scale distributed event. These tools have been adapted to be able to evaluate many of the platforms that participate in Virtual Flag (e.g. F-15, F-16, AWACS, ASOC, CRC, JSTARS, and JTAC). By using the same tools to collect data across multiple Virtual Flags, a large database of raw data will exist that can be used to evaluate training effectiveness of these kind of events. Further, over time we expect to achieve a sufficient sample size to permit detailed analysis of the impact of the training on mission performance and on proficiency. The consistent use of common instruments and protocols ensures that a comparable dataset representing the overall performance constructs of interest as well as unique training benefits and impacts accrued platform by platform, can be developed and validated over time. Finally, the data can be used to establish the training utility and readiness benefits associated with the integration of high fidelity and distributed simulation into continuation training for operational military personnel.

Acknowledgements

The authors would like to thank Peter Neubauer for his technical assistance with the mini-browser.

References

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

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.

Crane, P., Skinner, M., Best, C., Burchat, E., Gehr, S. E., Grabovac, M., Pongracic, H., Robbie, A., Zamba, M. (2006). Exercise Pacific Link: Coalition Distributed Mission Training Using Low-Cost Communications. Presented at: SimTecT 2006 - Simulation Conference and Exhibition, Melbourne Australia, June 2006.

Gehr, S. E., Schreiber, B., & Bennett, W., Jr. (2004). Within-Simulator Training Effectiveness Evaluation. Published in proceedings of The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2004), Orlando, Fl.

Gehr, S. E., Schurig, M., Jacobs, L., van der Pal, J., Bennett, Jr., W.; & Schreiber, B. (2005) "Assessing the Training Potential of MTDS in Exercise First Wave." In, *The Effectiveness of Modeling and Simulation – From Anecdotal to Substantive Evidence*. Meeting proceedings NATO Research and Technology Organization Modeling and Simulation Group RTO MP MSG 035. Neuilly-sur-Seine, France: RTO.

Smith, E., McIntyre, H., Gehr, S. E., Schurig, M., Symons, S., Schreiber, B., & Bennett, Jr., W. (2005). "Evaluating the Impacts of Mission Training via Distributed Simulation on Live Exercise Performance: Results from the UK/UK 'Red Skies' Study". In *The Effectiveness of Modeling and Simulation – From Anecdotal to Substantive Evidence*. Meeting proceedings NATO Research and Technology Organization Modeling and Simulation Group RTO MP MSG 035. Neuilly-sur-Seine, France: RTO.