Implementing an Automated Performance Assessment Capability for the Joint Training Special Event (JVTSE)

Elizabeth Biddle, Michael Keller
The Boeing Company
Orlando, FL
elizabeth.m.biddle@boeing.com,
michael.l.keller@boeing.com

Randy Pitz
The Boeing Company
St. Louis, MO
randy.pitz@boeing.com

Anthony Nixon
Boeing Australia Limited
Brisbane, QLD, Australia

anthony.d.nixon@boeing.com

ABSTRACT

The Joint Forces Command (JFCOM) demonstrated the Joint National Training Capability (JNTC) as part of a special event, called the Joint Virtual Training Special Event (JVTSE), during the 2004 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). JVTSE implemented a prototypical live-virtual-constructive (LVC) training event that was supported by participants from the government, industry, and academia. The assessment of large-scale, joint exercises is a laborious and challenging task. The current method employs human observers/trainers (O/Ts) monitoring the training audience's performance. Given the pace that these exercises are conducted and the large number of simultaneously occurring events, this assessment process is fraught with errors, including the omission of key events, and O/T bias. The need for automated performance assessment technologies to provide detailed and accurate reports of performance has been advocated (Cardinal et al., 2004; Meliza et al., 1992; Watz et al., 2003). This paper will target the design, development, and implementation of an automated performance assessment capability that; logs, analyzes, and presents data obtained from an LVC environment. It will provide information regarding the performance of the trainees as well as the training system. Additionally, this paper will describe how this system can be used to diagnose specific performance issues and make appropriate adjustments to improve human and/or system performance.

ABOUT THE AUTHORS

Elizabeth Biddle is an Engineer/Scientist at The Boeing Company located in Orlando, Florida. She is a Principal Investigator for advanced instructional and training research and development projects. Dr. Biddle earned a Ph.D. in Industrial Engineering and Management Systems from the University of Central Florida in 2001. She received the Modeling & Simulation Professional Certification by the Modeling & Simulation Professional Certification Commission in 2002 and the Modeling & Simulation Award in Training by the Defense Modeling & Simulation Office in 2001.

Michael Keller is a Senior Systems Analysts at The Boeing Company in Orlando, Florida. Mr. Keller is currently the Training Support Contract II Lot I Lead Engineer and is supporting the automated performance assessment research and development activities. Mr. Keller received an B.A. in Industrial Psychology from Michigan State University in 1966 and has graduate work at Ball State University 1970-71 and the University of Northern Colorado in 1974-75 in Psychology. Mr. Keller retired from the United States Air Force where he was dual rated as a Pilot/Navigator.

Randy Pitz is an Engineer/Scientist at The Boeing Company located in St. Louis, Missouri. Mr. Pitz is currently the Principle Investigator for desktop systems research and development projects. He received a B.S. in Computer Science from Michigan Technological University in 1992, and a M.S. in Computer Science from Washington University in 2004. He has over 12 years as a professional software designer and architect.

Anthony Nixon is the Technical Lead of the Systems Analysis Laboratory in Boeing Australia, and has almost 20 years experience as a software designer and architect. Anthony received a Bachelor of Applied Science (Computing) degree from the Queensland University of Technology in 1986 and since that time has worked for a variety of software companies in Australia and the US before joining Boeing Australia in 1999. His research interests include framework based development, visualisation architectures and simulation and modeling.

Implementing an Automated Performance Assessment Capability for the Joint Training Special Event (JVTSE)

Elizabeth Biddle, Michael Keller
The Boeing Company
Orlando, FL
elizabeth.m.biddle@boeing.com,
michael.l.keller@boeing.com

Randy Pitz
The Boeing Company
St. Louis, MO
randy.pitz@boeing.com

Anthony Nixon
Boeing Australia Limited
Brisbane, QLD, Australia

anthony.d.nixon@boeing.com

INTRODUCTION

The Joint Forces Command's Joint Virtual Training Special Event (JVTSE) presented at the 2004 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) demonstrated the feasibility of conducting a mutually supportive live-virtual-constructive (L-V-C) training event. The purpose of the JVTSE was to show how L-V-C training can support the JNTC's four pillars of effective training: (1) realistic combat training, (2) adaptive and credible forces, (3) common ground truth, and (4) high quality feedback. This paper will focus on the fourth pillar – the provision of high quality feedback.

The assessment of large-scale, joint exercises is a laborious and challenging task. Currently, joint exercises are evaluated by human observers/trainers (O/Ts) who monitor various aspects of the trainee's performance. The O/Ts are limited to the use of paper-based checklists enforcing subjective rating scales during their evaluations. Given the pace that these exercises are conducted and the large number of simultaneously occurring events, this assessment process is fraught with errors, including the omission of key events, and O/T bias. The need for automated performance assessment technologies to provide detailed and accurate reports of performance has been advocated (Cardinal et al., 2004; Meliza et al., 1992; Watz et al., 2003).

Automated performance measurement (Campbell et al., 2003; Schreiber et al., 2003) can be defined as the analysis of raw data obtained from a training environment to provide information to assist in the assessment of training objectives. This paper will target the design, development, and implementation of an automated performance assessment capability that; logs, analyzes, and presents data obtained from an L-V-C environment.

BACKGROUND

Joint Training Environment
The Department of Defense (DoD) has initiated Training Transformation (T2) to foster the development and implementation of innovative new training approaches. The goal is to prepare the Warfighters for the extremely volatile and constantly changing environment of today's military operations. T2 is supported by three identified capabilities, one of which is the Joint National Training Capability (JNTC). The JNTC's charter is to provide a joint context (e.g., environment, infrastructure) for both single service and multi-service training events. These events can be conducted in either horizontal (single echelon) or vertical (multi-echelon − e.g., strategic, operational, and/or tactical levels). To accomplish this task, the JNTC is facilitating the development of advanced technologies that support the previously mentioned four pillars of effective training.

The JVTSE was experiment conducted to investigate the linkage of distributed, L-V-C training assets to provide a realistic joint training environment for geographically distributed forces. The JVTSE integrated in excess of 20 platforms and systems on the I/ITSEC show floor as well several systems that were located off-site. The JVTSE highlighted the key capabilities enabled by an L-V-C distributed training environment. As the event's scenario was purposely created to showcase examples of the four pillars, the scenario was purely fictitious and did not employ current operational tactics, techniques or procedures.

The JVTSE scenario consisted of a series of tactical events woven together utilizing a wartime storyline. Each tactical event incorporated a specific mission emphasizing the capabilities of a specific platform. A Command and Control (C2) element provided the linkage between events and an overall joint context. The integration of these events resulted in a seamless flow between L-V-C assets. Data was gathered and logged from the network. It was then used to demonstrate the provision for high quality feedback.

Consistent with the fictitious nature of the scenario, the performance measures evaluated were derived based on an analysis of the events in the scenario and were not doctrinally sound.

High Quality Feedback

High quality feedback refers to the recommendations and lessons learned from the evaluation of specific joint performance outcomes and standards called out in joint concepts and doctrine. Additionally, high-quality feedback provides information regarding the effectiveness of joint training, consisting of data that can be used to assess the readiness of the participating forces, and diagnostics of the competencies that may require further refinement. This is consistent with Major General Gordon Nash's statement that the development of "better tools for collecting, analyzing, and cataloging lessons learned…will significantly improve training feedback and enhance the commander's ability to evaluate the readiness of his forces" (Nash, 2004). The General's comments emphasize the need for training evaluation tools and meaningful feedback.

Technologies, processes, procedures, and concepts that facilitate the accurate and efficient evaluation of joint (collective) training performance must support high quality feedback. Tools that capture the vast amounts of data generated in the network centric operational training environment enabled by the JNTC are mandatory. Training systems are currently collecting and passing billions of bytes of data to the instructor or analyst for classification and evaluation. However, the tools that enable an understanding of this data are non-existent. Without a means of making sense out of these voluminous data sets, the data is just noise—not high-quality feedback (Keeter, 2004).

It almost goes without saying that these tools must provide correct and timely data to the user (e.g., training audience, instructor, analyst) in a form that is easily understood. Often, useful feedback is not the raw data that is extracted directly from the training environment. Instead, the training environment data requires processing and transformation to provide valid and useable assessment of mission performance. Unfortunately the processing of raw data is both a time and labor-intensive procedure. To meet the vision of the JNTC, a mechanism or tool set that offers the ability to transform raw data into assessments of performance—in real-time—is evolving into a stated requirement. For the JVTSE, a simulation analysis tool was leveraged allowing the transformation of data

logged from the network into descriptive information of performance during the JVTSE scenario.

Joint Training Evaluation

The evaluation of joint and collective (e.g., single service) tasks is currently recognized as a laborious activity. Joint training evaluation, as it exists today, is conducted by subject matter experts and/or military analysts who observe the various aspects of the training event. This involves the parallel performance of various tasks by the trainees. The observers then subjectively determine the trainee's mastery of the training objectives identified for the training event (Campbell et al., 2003; Holden, Throne, & Sterling, 2001).

The data recording capabilities provided by today's state-of-the-art training technologies further exacerbates the ability to observe all of the activities involved with the training event. The complexity and abundance of information diminishes the overall effectiveness and accuracy of the evaluations presented during the AAR, which is actually the portion of the training exercise in which learning most often takes place (Meliza et al., 1992). Additionally, the ability to report on the effectiveness of these environments is subsequently reduced (Stephens, 2004).

For quite some time, researchers (Cardinal et al., 2004; Meliza et al., 1992; Watz et al., 2003) have advocated the need for automated, objective measures to provide the instructor with detailed and accurate information on collective performance. The intent is to improve training feedback, (such as in AAR), and to evaluate the effectiveness of training. The term automated performance measurement (Campbell et al., 2003; Schreiber et al., 2003) has therefore been coined and refers to the statistical analysis of raw data (e.g., obtained from a virtual training environment or live training instrumentation) to provide feedback on a trainee's proficiency with respect to a particular training objective or competency.

Military operations involve simultaneous performance of task-related and team-related (e.g., communication, coordination, etc.) processes. Both processes involve the synchronous execution of lower-level behavioral processes and higher-level cognitive processes (Oser et al., 1999). The former refers to the procedural skills described by standard operating procedures (SOPs), operation orders, standards, and doctrine. For instance, departing at the scheduled and prebriefed time or opening a secure communications net with the proper frequency and clock setting are examples of behavioral

processes. The latter refers to the ability to integrate knowledge with the appropriate level of situational awareness to make correct and timely decisions. An example of a cognitive process is decision-making, such as attacking opposing forces when you do not posses the optimum and/or required number of combat assets.

Behavioral aspects of individual task performance are evaluated in terms of measures of performance (MOPs) and/or measures of effectiveness (MOEs). MOPs evaluate the execution of processes, while MOEs evaluate the outcome of processes. MOPs and MOEs provide a measure of what happened (Campbell et al., 2003) and are evaluated objectively through the recording of events in the training environment (by an observer or computer data logger) that signify the execution of a procedure (e.g., keypress, data entry, etc.) or an outcome (e.g., number legitimate kills, etc.). Automation of MOPs and MOEs can be accomplished by associating recorded data with performance standards (Campbell et al., 2003).

The evaluation of cognitive processes is more complex because it involves the determination of why specific behaviors were executed—a difficult task that has been addressed by many researchers. Inherently, cognitive process evaluation is not easily automated. The multiple challenges in accurately assessing cognitive processes often result in this important aspect of

performance being ignored during AARs (Campbell et al., 2003). Cognitive processes, for the most part, are not directly observable by a human or computerized data logger. Rather, they are evaluated by inferring relationships between the decision-maker's observable behaviors and his cognitive processes (e.g., verbal and nonverbal communications between team members, implied signals, etc.). Humans are required to perform these observations and apply subjective evaluations because the relationships between these types of behaviors and cognitive processes are nonlinear.

As training tasks grow more complex (increased joint force training) and the size of the units being trained increase, the more difficult – or impossible – this type of evaluation becomes. Technology to support the evaluation of cognitive processes for collective training is required to ensure improvement in training effectiveness. Methods and technologies for objectively measuring cognitive processes will improve the precision and accuracy (Watz et al., 2003) of their evaluation. The notion of automated performance assessment – the collection and translation of raw data through pre-determined algorithms to provide meaningful performance assessment feedback that enhances traditional performance measurement methods – has been advocated by many (Campbell et al., 2003; Cardinal et al., 2004; Schreiber, Watz, Bennett, & Portrey, 2003; Thorne & Holden, 2003).
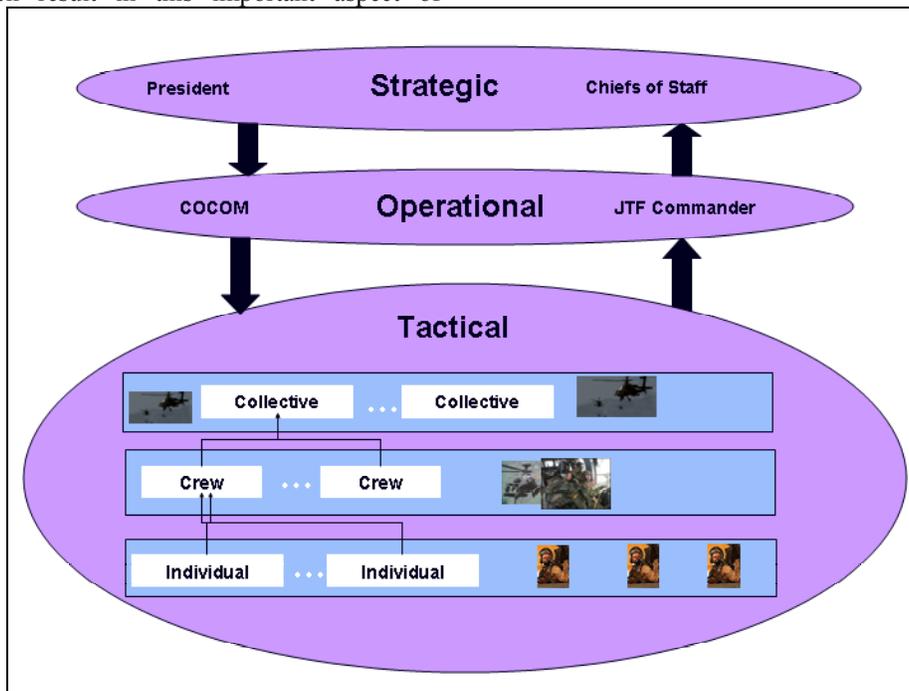
Figure 1. Collaborative Training Levels

APPROACH

Effective joint training requires a linkage between the top-level mission objectives and the lower-level collective/unit and individual training objectives (Figure 1). JFCOM is addressing this issue through the development of joint training tasks that decompose joint training articles from the high-level (operational, staff) tasks down to the supporting tasks at the service level (Vesely, 2005). A linkage between top-level mission objectives and lower-level human and machine-system performance objectives will facilitate the root cause analysis of specific mission outcomes.

The automated performance metrics demonstration sought to show the evaluation of performance across the operational and tactical levels in an effort to improve understanding of the JVTSE's scenario performance. As described in the previous section, the automated performance assessment tools employ data collected from the training environment to compute pre-defined metrics. Therefore, metrics for the joint and unit level were derived from the scenario information available to the authors prior to I/ITSEC. The following paragraphs describe the process of specifying these metrics.

Initial Metric Development

The performance metrics were derived from the scenario. In order to allow for sufficient time to implement the automated performance metrics for the demonstration, the task of developing the metrics started immediately after the first draft of the scenario was released. The scenario went through several iterations that directly impacted the planned metrics. As a result, the metrics specified for the JVTSE also went through several iterations. Additionally, specific target platform entities and their exact locations were not documented in some cases. The overall result was that some of the metrics that required specific target tracking data were not implemented. Also, the scenario was originally developed on an event-based rather than time-driven sequence of events, and therefore, the metrics developed did not include time on target (TOT) or other time-related standards. (Later, the scenario adopted a time-based sequence of events.)

The original set of metrics defined: operational level outcome measures (evaluation of mission success), operational level process measures (evaluation of actions performed at a joint level), and tactical level task and teamwork process measures (evaluation of a unit's task and teamwork actions).

Territorial Integrity Restored	
Definition	No live/functional Tularian entities in the territory of Colusa.
Method	Identify state of Tularian entities within the boundary of Colusa at the conclusion of the scenario.
Score	Achieved or not achieved (P/F).
Data Requirements	OPFOR entity state (position and live/dead status) – recorded at the conclusion of the scenario. Location (coordinates) of the OPFOR entity within the boundaries of Colusa so a determination can be made as to whether the entity is within the bounds of Colusa.

Figure 2. Operational Level Outcome Measure Example

Operational Level Outcome Measures

The Operational Level Outcome Metrics were developed to support the assessment of the accomplishment of mission objectives. As stated previously, the JVTSE was mission was fictitious and developed for the purposes of conducting an experiment to integrate L-V-C assets to provide a distributed training capability. These mission objectives were to maintain air superiority and restore territorial integrity. One of the Operational Level Outcome Measures is provided in Figure 2. (For the JVTSE scenario, Tularia was the name of the invading country, and Colusa was the defending country.)

Was the CAS Mission completed with no fratricide?	
Definition	Metric M 1.6 from the Conduct CAS Extended Joint Task Article 3.2.2.
Method	If BLUFOR entity's weapon hits another BLUFOR entity, then evaluated as no, else yes.
Data Requirements	BLUFOR entity weapon launch and detonation.

Figure 3. Operational Level Process Metric

Operational Level Process Measures

The Operational Level Process Measures were developed to provide a means of evaluating the proficiency of operational level tasks. The authors identified relevant metrics from the Execution Task Measures defined by the *Extended Joint Task Articles (JTAs)* for JTA 3.2.2 Conduct Close Air Support (CAS). Per the Extended JTA, each of the measures is evaluated as Yes/No – meaning a specific task was or

was not performed. Figure 3 shows one of the Operational Level Process Metrics that was developed.

Tactical Level Measures

The Tactical Level Measures were developed to identify both task and teamwork processes. These measures are intended to assist the unit commander level with evaluated performance of his unit. The unit's performance evaluation can be used to provide an understanding of effects observed at the operational level.

Unit tasks are typically evaluated with a paper-based checklist that the instructor/observer completes indicating "Go/No Go" for the performance of a specific task. The specification of the unit level missions, which were referred to as the scenario's events, were not at the granularity of individual performance as that was outside of the JVTSE focus. Therefore, the authors used creative license in identifying event tasks and metrics. An example of one Unit Task Metric is provided in Figure 4.

Task: Conduct F/A-16 Airfield Strike
1. Sub-task 1: Acquire Target (Is target painted by the F/A-16's radar?)
    a. Data requirement: F/A-16: radar (determine if target is being painted by the radar)
2. Sub-Task 2: Lock Target (Does one of the F/A-16s have the target designated?)
    a. Data requirement: F/A-16 indication that target is acquired and locked
3. Sub-Task 3: Shoot target (F/A-16 employs weapon)
    a. Data Requirement: F/A-16 weapon launch
4. Sub-Task 4: Kill target (weapon employed on correct target)
    a. Data requirement: F/A-16 weapon detonation
5. Sub-Task 5: Egress (fly back to designated rendezvous point)
    a. Data requirement: F/A-16 heading change toward rendezvous point

---

Figure 4. Unit Task Metric Example

As described earlier, teamwork process (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Spiker, Silverman, Tourville, & Nullmeyer, 1998) refers to those activities performed when working as a team, such as: communication, coordination, decision-making, situational awareness, task management, and

mission evaluation. Teamwork processes are typically evaluated based on subjective evaluation, as these types of processes are difficult to define. Behaviorally Anchored Rating Scales (BARS) are subjective evaluation instruments that define behaviors for each scale value to improve inter-rater reliability. For the Unit Teamwork Metric, behaviors related to Task Management were defined with the goal of using scripted rules to evaluate the behavior definition based on data obtained from the training environment. Figure 5 shows the description of the Navigation Accuracy metric.

Navigation Accuracy	
Definition	Degree to which units reached pre-planned waypoints.
Method	Compare the position of unit platform 1 and unit platform 2 to the proper waypoints at the time that these waypoints are over flown or passed.
Score	Rated on a 1-5 scale that evaluates the difference between the unit's actual position and the unit's preplanned position. (e.g., 5 = actual and preplanned position are the same AND planned and actual time over position are the same, 4 = actual position deviates from planned position by less than 100', etc.)
Data Requirements	Platform position and waypoint positions.

Figure 5. Unit Teamwork Metric Example

Metric Refinement

These metrics were shared with the JVTSE scenario-working group and associated JFCOM personnel per their request. Feedback received from the metric review was mixed. Concerns that the operational level outcome measures would not be accurate since the mission was not doctrinally based were voiced. Additionally, the lack of time on target criteria identified the difficulties in locating data related to a specific event in a DIS data log (e.g., data analysis could not be limited to specific time intervals in which to evaluate performance). As collective training is often event-based, this lesson learned was valuable in identifying factors that need to be considered in evaluating metrics in a distributed training environment. Again, many of the specific target entities and exact locations were not documented. As a result, the operational level outcome metrics were deleted and the remaining metrics were reduced. Figure 6 provides the

final list of metrics that were implemented for the JVTSE demonstration.

JVTSE PERFORMANCE ASSESSMENT

Data analysis and recording tools were used to automate the evaluation of the metrics presented in the previous section. The following briefly describes the data logging and analysis tools, and their integration for the evaluation of the metrics defined in Figure 6. This implementation was developed to support a post-event analysis of the JVTSE scenario to support a high-level AAR.

Average Range Destroyed: Calculate the range between the BLUFOR and Target at the time the BLUFOR platform destroyed the target (time at which missile hit the target). If they are equal, evaluated as yes, else no.
Number of Targets Destroyed: The number of targets reported as killed divided by the total number of intended targets.
Fratricide Occurrence: How many blue-on-blue weapons were fired.
Number of Communications: Calculate the number of communications made between BLUFOR entity and AWACS to provide an indication of the coordination between AWACS and BLUFOR entity.
Correct Target Struck: Determines if the BLUFOR entity's weapon hit the intended target. If they are equal, evaluated as yes, else no.

Figure 6. Metrics Implemented for JVTSE

Tools

Data logging and data analysis used were used together to provide an automated performance assessment capability. These tools are described in the following.

BASE

BASE is a simulation analysis tool that has been created by the Systems Analysis Laboratory (SAL) in Boeing Australia (www.boeing.com.au/SAL/sal.htm) for the purpose of supporting the analytical requirements of simulation-based operations analysis. The overall aim of BASE was to reduce the requirements placed on the analyst in manipulating data, allowing them to focus on the key insights to be gained from the simulation results. By being able to "read" the output of virtually any piece of simulation software, it also placed less stress on the ability of team members to be specialists in one particular tool or another, allowing a more effective approach to problem solving not constrained by one or more tools.

BASE consists of the following applications:
- BASE Command - An intelligent file system explorer that automates many of the tasks associated with organizing and recording the output from simulation applications.
- BASE Debrief Realtime – Analyses the data produced by a single run of a simulation, possibly in real-time as the simulation is running.
- BASE Analyst – Analyses and visualizes simulation data in a number of ways, and allows the comparison of key metrics across multiple simulation runs.

The automated performance metrics were implemented with BASE Analyst.

JVTSE Data Collection

The MAK Technologies, Inc. Data Logger was used to record the simulation data during the JVTSE demonstration. BASE has the capability to read these log files using the MAK VRLink runtime. The simulation data was also logged using a prototype Boeing logger, for after-action review (AAR) or playback capability. Data were logged for each of the four iterations of the JVTSE, and the rehearsal held on the I/ITSEC show floor. The files were saved and read by the BASE Analyst tool.

Metric Analysis

The pre-defined analyses were defined for the five metrics identified in Figure 6 using the BASE Analyst analysis specification features. BASE is built upon a software framework designed to facilitate the processing and analysis of simulation data. The key elements of this framework are shown in Figure 8.

	Attacking Force	Target Force	Quantity	Percentage	Quantity Impacted Entity	Percentage	Accuracy
1	Friendly	Opposing	3821	97.89905201	18	78.26086957	0.471081
2	Opposing	Friendly	22	0.563668972	3	13.04347826	13.6364
3	Opposing	Opposing	29	0.743018191	2	8.695652174	6.89655

Figure 7. Fratricide Occurrence Metric - Table
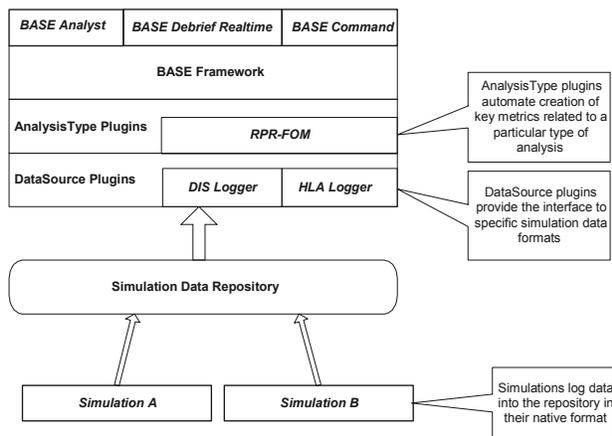
**Figure 8. BASE Framework Architecture**

The two key areas in which the BASE Framework can be extended are in the areas of interfacing to simulation data schemas (via the *DataSource* abstraction) and the area of automating the extraction of key measures of effectiveness (via the *AnalysisType* abstraction).

The *DataSource* acts as a source for the visualisation pipeline, creating instances of *Metric* that represent the fundamental data provided by the simulation. This allows BASE itself to be simulation independent, through the development of *DataSource* based simulation specific plugins.

Instances of class *AnalysisType* encapsulate the knowledge required to create an analysis pipeline to extract measures-of-effectiveness from source metrics. Since analysis types are only loosely coupled to input metrics, reuse is promoted.

For the JVTSE exercise, the *DataSource* plugin used was a DIS based one, and the *AnalysisType* plugin was one based on the standard RPR-FOM/DIS data model.

A representation of the process that occurs to create the metric shown in Figure 7 is shown in Figure 9.
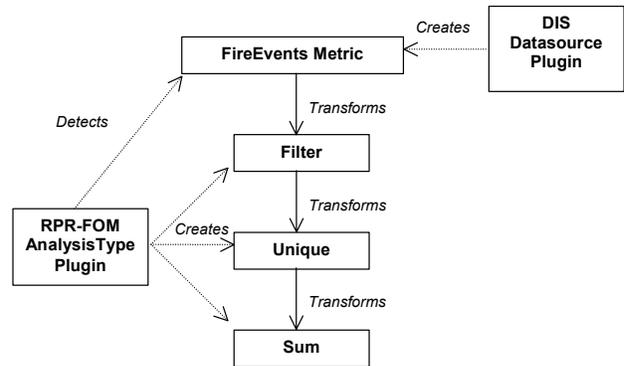
Figure 9. Example Analysis Pipeline

As shown in Figure 9, the DIS Logger Datasource plugin detects the presence of DIS based data and automatically creates a number of metrics to represent the low-level data able to be extracted from a DIS data stream. In this diagram, only the "FireEvents Metric" is shown, which corresponds to information extracted from the DIS FireEvent PDU. The datasource in question actually also creates a large number of other metrics, each corresponding closely to low-level data published by a DIS simulation.

Once the low-level metrics have been created by one (or more) DataSources, the installed AnalysisTypes examine the metrics which exist and create higher-level metrics that transform the low-level data into meaningful measures of effectiveness. In the case in point, the "Force on force munitions expended" metric is created by filtering the FireEvents metric, then selecting unique records and then summating those records, creating the final metric seen in Figure 7.

This architecture is particularly suited to the automation of performance assessment since it readily supports the ability to:

1. Process data from several data formats in the one environment. For example, a DataSource can be created to process instructor assessment data, along with a DataSource that processes DIS based truth data.
2. AnalysisType plugins can be created that correlate and process these diverse datasets to provide an easy to use assessment of overall outcome.

The metrics were available for presentation in tabular and various graphical formats. For example, Figure 7 displays the quantity of munitions exchanged between various configurations of attacking and target forces, as well as the number of successful impacts. Note that

there is no data for Attacking Force: Friendly – Target Force: Friendly, this indicates that there were no instances of friendly on friendly fire. As such, there were no instances of fratricide for this trial. Figure 10 presents the data in a network diagram.
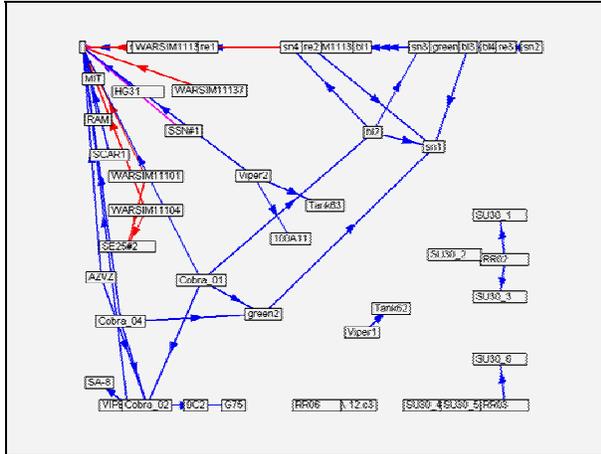
Figure 10. Fratricide Occurrence Metric – Network Diagram

LESSONS LEARNED

The evaluation of joint training performance is much more than just a technological challenge. Performance evaluation requires prior consideration and the specific identification of the areas in which feedback is desired. The following describe the lessons learned from the process of implementing an automated performance metric demonstration in support of the JVTSE.

Target Specification

For the majority of the events, the engineers for the various systems defined the entities that comprised the targets without documenting the specific entity types comprising the target and their various locations (e.g., coordinates) within the simulated environment. However, the authors assumed that the target definitions and locations defined in the scenario documentation were what should be expected in the actual scenario. If this had been an actual training event, rather than an experiment addressing L-V-C training, this most likely would not have occurred. However, it is an important reminder that in order to develop metrics that evaluated whether or not the correct targets were destroyed, information regarding the specific entity (ies) that comprised a target and its initial location was needed.

Our solution to this problem for JVTSE was to only evaluate whether or not the correct target was destroyed for the first two events because the entity type and

location data was available for the targets involved in these two events. We recommend that in future events that the target entity types and locations be precisely documented. Additionally, we recommend that the documentation be strictly adhered to when implementing the scenarios.

Time Specification

During the JVTSE Final Planning Conference, the scenario committee decided that the scenario would be event-based vice time-based. (Note: This was approach was adopted in support of the demonstration objectives, so would likely not be an issue in an actual training event.) The inability to define the events by time period of occurrence made it difficult to filter network traffic as elapsed time could have been used as a marker in filtering the data. A time-based approach was adopted for the actual performance of the scenario at I/ITSEC, but this change occurred too late for modifications to the metrics.

We addressed this problem by avoiding the use of time-on-target metrics. Additionally, we utilized only the first two events since it was possible to identify them without a specific time window to filter the data. We recommend future events define and use time-on-target criteria for the specific events that are to be evaluated.

Scenario Development

The final scenario was not completed until the integration and testing period. This did not leave sufficient time for the development of the metrics after the completion of the scenario. Rather, it was necessary to specify the metrics during the scenario development stage. Therefore, only high-level metrics (e.g., whether or not fratricide occurred, correct target was struck, and BLUFOR range from target when destroyed) that were applicable to a range of joint level exercises were specified.

As performance assessment is used to evaluate training objectives, it is critical that the training objectives be defined prior to the development of the metrics. Therefore, we recommend that the specification of mission/training objectives and criteria for success be a high-priority in the scenario development process.

Entity Enumeration

Many of the entities involved in the scenario were non-standard and not defined in the enumeration list. The packets for these entities lacked documentation within most of the entity state fields. The evaluation of

performance requires the ability to associate data from the network to the entities (platform, human, weapon, etc.) of interest. For the JVTSE demonstration, the use of BASE enabled us to label some of the unknown data types at a high-level (e.g., land or surface platform). However, we recommend the use of standard entity types or the documentation of new entity types according to published standards.

CONCLUSION

The use of L-V-C technologies provides a data rich environment for extracting useful information with respect to training performance. Tools exist to assist in the evaluation of training outcomes, but work is required to develop tools that better evaluate trainee processes.

Logging the JVTSE data not only supported the performance assessment goals, but also gave the JNTC the ability to troubleshoot problems which are inherent to these large exercises. An exciting side affect of logging data for human performance assessment is to provide system validation and network troubleshooting, or systems performance assessment.

The implementation of performance metrics for the JVTSE demonstrated the use of automated metrics to evaluate task work and teamwork process at the operational and tactical level. The process of developing and implementing these metrics for the JVTSE was beneficial in gaining insight on ways to better implement automated performance assessment technologies.

ACKNOWLEDGEMENTS

REFERENCES

Campbell, C.H., Throne, M.H., Black, B.A., & Lickteig, C.W. (2003). *Research observation and lessons learned for the Future Combat Systems* (Research Product 2003-04). US Army Research Institute for the Behavioral and Social Sciences.

Cannon-Bowers, J.A., Tannenbaum, S.I., Salas, E., & Volpe, C.E. (1995). *Defining team competencies: Implications for training requirements and strategies*. In, R. Guzzo & E. Salas (Eds.), Team effectiveness and decision making in organizations. San Francisco: Jossey Bass.

Cardinal, C., McTigue, K., Nguyen, L., Bergondy, M., & Merket, D. (2004). *Initial development and continuing evolution of the performance measurement base object model (BOM) for the Navy Aviation Simulator Master Plan (NASMP)*. In, Proceedings to the 2004 Spring Simulation Interoperability Workshop.

Holden, W.T., Throne, M.H., & Sterling, B.S. (2001). *Prototype Automated Measures of Command and Staff Performance* (Research Report 1779). Alexandria, VA: US ARI for the Behavioral and Social Sciences.

Keeter, H.C. (2004) *Urban operations challenge shows limits of U.S., Allied ISR Capability*. Sea Power, May, 14-15.

Meliza, L.L., Bessemer, D.W., Burnside, B.L., & Schlecter, T.M. (1992). *Platoon-level after action review aids in the SIMNET unit performance assessment system* (Technical Report 956). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Science.

Nash, G., Maj. Gen. (2004). *Transcript of the House Armed Services Committee Subcommittee on Terrorism, Unconventional Threats & Capabilities and Subcommittee on Readiness on the Joint National Training Capability* (March 18). Available at: http://www.jfcom.mil/newslink/storyarchive/2004/sp031804a.htm.

Oser, R.L., Cannon-Bowers, J.A., Salas, E., & Dwyer, D.J. (1999). *Enhancing human performance in technology rich environments: Guidelines for scenario-based training*. Human/Technology Interaction in Complex Systems, Vol. 9 (pp. 175-202). JAI Press, Inc.

Schreiber, B.T., Watz, E., Bennett, W., & Portrey, A.M. (2003). *Development of a distributed mission training automated performance tracking system*. In, Proceeding to the 2003 Behavior Representation in Modeling & Simulation Conference. Downloaded from the World Wide Web on April 19, 2004: http://www.sisostds.org

Spiker, V.A., Silverman, D.R., Tourville, S.J., & Nullmeyer, R. (1998). *Tactial Team Resource Management Effects on Combat Mission Training Performance* (AL/HR-TR-1997-0137). Mesa, AZ: Air Force Research Laboratory Human Effectiveness Directorate Warfighter Training Research Division.

Stephens, H. (2004). *DoD Needs Better Ways to Assess Effects of Networked Warfighters*. GAO (July 2).

Watz, A., Schreiber, B.T., Keck, L., McCall, J.M., & Bennett, W. (2003). *Performance measurement*

challenges in distributed mission operations environments. In, Proceedings to 2003 Fall Simulation Interoperability Workshop. Downloaded from the World Wide Web on April 5, 2004: http://www.sisostds.org

Throne, M. & Holden, W. (2003). *Prototype automated performance measures for future battle staffs*. 2003 Interservice/Industry Simulation, Education, and Training Conference Proceedings (Dec 1-4, 2003).

Watz et al., 2003

Vesely, D. CDR. JTA / JC MTG Development and JTRAT. 4 March 2005.