A Smart Approach for After Action Review Visualization and Analysis

Jennifer K. Phillips, Allison Hancock, Marc Pfahler
Cognitive Performance Group
Portsmouth, VA
{jenni, allison, marc}
@cognitiveperformancegroup.com

Marcus Mainz, Luke Cardelli, Breck Perry
Covan Group
Stafford, VA
{marcus.mainz, luke.cardelli, breck.perry}
@covangroup.com

Christopher Young, Richard Schaffer
Lockheed Martin
Boston, MA
{christopher.a.young, richard.l.schaffer}
@lmco.com

Gwen Campbell, Audrey Zlatkin, Costas D. Koufogazos
Design Interactive, Inc.
Orlando, FL
{gwen.campbell, audrey.zlatkin, costas.koufogazos}
@designinteractive.net

ABSTRACT

Live field training is essential to the preparation of Marines and Marine units. No other training environment can approximate real-world combat conditions including time pressure, the fog and friction of war, and in the case of force-on-force training, a dynamic adversary. The subsequent after-action review (AAR) is where most learning takes place. An effective AAR facilitates reflection by providing feedback on individual and unit performance so that strengths can be sustained and deficiencies, corrected. The purpose of the Office of Naval Research's Streamlined Marine After-Action Review Tool for Visualization (SMART-Viz) research and development effort is to optimize learning and the development of expertise with data-driven feedback on unit performance.

The SMART-Viz research program uses high-fidelity, integrated, intelligent AAR tools to capture objective data about what occurred rather than subjective perceptions, and visually reproduce the events of the battle across the warfighting functions. The SMART-Viz tools rapidly synchronize data from instrumented sources and observers to provide information about tactical events paired with cognitive and behavioral performance assessments of decision making. Individual and unit performance measures are displayed in an analysis dashboard combining a visual representation of ground truth on the battlespace, battle outcomes, and performance measure graphics. Expected results include increases in the volume and objectivity of data available to AAR facilitators and utility of the data for pinpointing performance deficiencies and delivering visual battle replay feedback that promotes comprehension and learning. This paper describes the measures forming the foundation of the AAR analysis and early results drawn from Marine Corps exercises including battle outcomes and measures of decision making. Specifically, outcome measures of survivability and lethality were collected via data logs and sensors. In addition, quality measures of commander and staff decision making from observer ratings were collected and tested.

ABOUT THE AUTHORS

Jennifer K. Phillips is the Chief Executive Officer and a Senior Scientist at the Cognitive Performance Group. Her research interests include skill acquisition, cognitive performance improvement, and the nature of expertise. Ms. Phillips has over 25 years of experience modeling performance across the levels of proficiency, designing learning solutions including decision-centered training scenarios and facilitation techniques, and developing metrics for cognition and decision making.

Dr. Allison Hancock is a Senior Scientist at the Cognitive Performance Group studying operational military research and human performance assessment. Dr. Hancock has a Ph.D. in Educational Psychology with specialization in Sport and Exercise Psychology and Measurement and Statistics, and a M.A. in Applied Experimental Psychology. She has over ten years of experience leading research projects with distributed teams for the DoD in the areas of human performance, rapid training development, cognitive skills training, program evaluation, and assessment.

Marc Pfahler is a Research Scientist at the Cognitive Performance Group. His primary research interests include human trust factors, multimodal biometric systems, and human interactions with autonomous systems. His research has been implemented in Air Force training techniques for machine automation of visual searches in complex environments. Mr. Pfahler has extensive experience in human-machine interactions and in data analysis, including questionnaires, quantification, and statistical analysis.

Marcus Mainz, USMC (Retired), is the Chief Executive Officer of Covan Group, LLC. Marcus retired from the Marine Corps after 20 years of commissioned service. Marcus's positions during his career include Future Operations Officer and Infantry Battalion Commander with the 2nd Marine Division; MAGTF Planner for II MEF and 2d MEB; Infantry Battalion Operations Officer and Infantry Company Commander with the 3rd Battalion, 7th Marines in Iraq; Expeditionary Warfare School Instructor; and Infantry Officer's Course and The Basic School Instructor. During Command, Marcus's battalion worked with the Office of Naval Research to develop the Tactical Decision Kit.

Luke Cardelli, USMC (Veteran), is the Director of Multi-Media Operations at Covan Group, LLC. Luke served in the Marine Corps for 5 years and achieved the rank of Sergeant. In his active duty time he served as a Squad Leader in the 3rd Battalion, 5th Marine Regiment, and supported the Marine Corps Warfighting Laboratory's (MCWL's) Sea Dragon Experiment where he executed Adversarial Information Operations. Upon leaving the Marine Corps, Luke worked an additional 3 years at the MCWL in the same capacity. With Covan Group, Luke currently supports the data collection process of both live and simulated experiments and training exercises.

Dr. Breck Perry, USMC (Retired), is the Director of Operations for Covan Group, LLC and a retired Marine infantry officer. While in the Marines, Breck served in various command and staff positions across the globe in support of combat, humanitarian, and crisis response operations. As a Captain, Breck discovered his passion for teaching and writing, and served for six years as a non-resident Expeditionary Warfare School and Command and Staff College adjunct faculty instructor, earning the Thomas S. Jones Instructor of the Year Award in 2015, 2016, and 2020. Upon retirement in the spring of 2022, Breck completed his Ph.D. in Educational Leadership and joined Covan Group, LLC.

Christopher Young is a staff software engineer at Lockheed Martin Rotary and Mission Systems (RMS). and has been the lead developer and project engineer on several US Marine Corps and Office of Naval Research programs for Lockheed Martin's Advanced Simulation Center. Since 2014, Chris has been the lead developer of the Interactive Tactical Decision Game (ITDG) application and the related Analysis Dashboard. Mr. Young holds a Bachelor of Science degree in Aerospace Engineering from Boston University.

Richard Schaffer is a Lockheed Martin Fellow and Principal Investigator at Lockheed Martin Rotary and Mission Systems (RMS). He leads the Human Immersive Simulation Lab at RMS's Advanced Simulation Centers. Richard received his S.B. degree from the Massachusetts Institute of Technology and has over 35 years of experience in modeling and simulation research and development. His areas of research have included distributed simulation, environment modeling, immersive simulation, and augmented reality. In 2010 he received the NTSA's lifetime achievement award.

Dr. Gwendolyn Campbell is a Portfolio Manager at Design Interactive. Gwen holds a Doctorate in Industrial and Systems Engineering from the Georgia Institute of Technology and has over seventeen years of industry experience in human performance and human systems integration, simulation, applied statistics, operations research, lean six sigma, and project management and seven years teaching experience within industry and academia. Her most recent work focuses on the strategic creation of training tools for competency development in military environments, accelerating instructor mastery, and the evaluation of virtual and augmented reality training systems.

Dr. Audrey Zlatkin is a Senior Research Associate I at Design Interactive. She has over 8 years of experience in human factors, cognitive psychology, and human-systems integration with a focus on adaptive training and decision support technology for operational environments. Before joining DI, Audrey was a research associate at the University of Central Florida, with her research primarily focusing on categorization and decision-making, with the goal of improving human performance within operational environments.

Costas D. Koufogazos is a Research Associate II at Design Interactive in Orlando, FL. He holds an M.S. Degree in Human Factors from Embry-Riddle Aeronautical University and a B.S. Degree in Health Sciences from the University of Central Florida. Costas utilizes his Human Factors research and writing skills to support multiple projects. He evaluates user needs and translates customer requirements into user-centered specifications, tracks and manages Agile development needs, conducts analyses, and prepares technical reports.

A Smart Approach for After Action Review Visualization and Analysis

Jennifer K. Phillips, Allison Hancock, Marc Pfahler
Cognitive Performance Group
Portsmouth, VA
{jenni, allison, marc}
@cognitiveperformancegroup.com

Marcus Mainz, Luke Cardelli, Breck Perry
Covan Group
Stafford, VA
{marcus.mainz, luke.cardelli, breck.perry}
@covangroup.com

Christopher Young, Richard Schaffer
Lockheed Martin
Boston, MA
{christopher.a.young, richard.l.schaffer}
@lmco.com

Gwen Campbell, Audrey Zlatkin, Costas D. Koufogazos
Design Interactive, Inc.
Orlando, FL
{gwen.campbell, audrey.zlatkin, costas.koufogazos}
@designinteractive.net

INTRODUCTION

Live force-on-force training is essential to the preparation of Marines and Marine units. No other training environment can approximate real-world combat conditions including time pressure, the fog and friction of war, and a dynamic adversary. The first Marine Air-Ground Task Force (MAGTF) Warfighting Exercise (MWX) took place in California in 2019 as a means to produce decision-making "reps and sets" against an intelligent adversary, which more closely approximate the conditions in war. Similarly, formal schools like The Basic School (TBS) and the Infantry Officer Course are implementing live free play force-on-force culminating exercises.

The after action review (AAR) following a training exercise is where the foremost learning takes place (Kaliner, 2013). Thus, an effective AAR is essential to gain the full benefit of costly live training exercises. Such an AAR facilitates reflection on performance to provide feedback on individual and unit performance so that strengths can be sustained and deficiencies, corrected (Department of the Army, 1993). To deliver an effective AAR requires the facilitator to reproduce the events of the battle armed with accurate information about what occurred and objective data about individual and unit performance.

The Marine Corps requires an efficient, data-driven approach to developing and conducting exercise AARs. The fog of the exercise, friction of command, and general uncertainty mean that ground truth is often unknown. Current Marine Corps practice relies heavily on the judgment of subject-matter experts (SMEs) to report exercise events, produce AAR products, and provide feedback to the exercised units. In all, nearly 100 active duty and civilian personnel work 24-hour operations to run the battle and produce a meaningful AAR for exercised units. Subjective perceptions of events from different vantage points are combined to attempt to identify what occurred.

Furthermore, reliance on subjective SME perceptions and feedback without objective data or a method to share it means that individual units benefit from the SME insights, but the value stops there. Commanders and units are unable to see how they stack up against their peers. Performance trends across units cannot be captured. Impacts of previous training investments cannot be gauged. Most critically, participants are not assessed against the same standards because each SME applies their unique experience and bias to their feedback. The inconsistent delivery of critical feedback hinders development of decision-making expertise. The introduction of objective measures of performance in force-on-force exercises would substantially enhance the learning that occurs during and beyond the AARs.

Project Objectives

The purpose of this research and development effort is to optimize learning and the development of expertise in force-on-force exercises with data-driven AARs. More specifically, our goals are to (1) identify performance data and measures that will meaningfully contribute to decision-centered AARs, (2) support the efficient collection and integration of the performance measures, and (3) produce visualizations and analytics with intelligent AAR tools that can be manipulated by facilitators and instructors. To achieve these goals, the Streamlined Marine After Action Review Tool-Visualization (SMART-Viz) program is comprised of three integrated development thrusts. The first thrust is the identification and validation of performance measures. The performance measures are the primary topic addressed in this paper. The second thrust is the Analysis Dashboard, which ingests data from disparate sources to create a visualization of the battlefield, ground truth locations of units and events, and analytics associated with events or the entire battle. The Analysis Dashboard builds on Interactive Tactical Decision Game (ITDG) technology (Young et al., 2019). The third thrust is the Field Assessment System, a tablet-based application that supports observer/controller (O/C) performance ratings and event markings in the field. The Field Assessment System builds on the Interactive Military Instructor Training and Assessment TEchnology (IMITATE) tool (Campbell et al., 2021). The SMART-Viz performance measures and AAR tools are anticipated to transition to the Marine Corps with unlimited rights and integrate with other technologies being developed by the Marine Corps (e.g., exercise design tools). In contrast to other AAR visualization systems, the tools incorporate an extremely broad range of data sources, support a diverse range of exercise types, and will enable administration, tailoring, and use by Marine Corps and government civilian personnel.

PERFORMANCE DATA AND MEASURES

The first year of the project focused on defining data and performance measures that support learning, are agnostic to the scenario, and have relevance to force-on-force exercise environments across the live, virtual, and constructive continuum. Further, the data and measures must support learning for echelons from small unit to division. The performance measures identified consist of behavioral, cognitive, outcome, and Observe-Orient-Decide-Act (OODA; Hammond, 2018) decision-making measures.

Behavioral Data and Measures

Behavioral measures reflect actions of units such as their movements, the time to conduct a task, or rounds fired. Behavioral measures are often paired with simple but informative data about the exercise, such as locations of events like air-to-ground or indirect fire attacks, and the effects. Figure 1 illustrates the Analysis Dashboard visualization of behavioral data using color-coded "+" symbols denoting the type of event (the interior color) and the initiator of the event (the external color). Selecting an event with a mouse click produces data about the type, time, units involved, and adjudication of the event. These data are integrated in the Analysis Dashboard to produce a visualization of ground truth for the AAR facilitator.

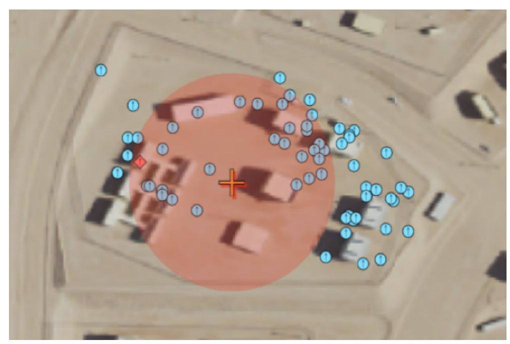
Figure 1. Analysis Dashboard Visualization of Key Exercise Events

Calculation of the measures requires reliable access to data. The challenge is to create a standardized approach to capturing the data across the variety of exercise contexts, despite differing resource availability and exercise control processes. In some smaller exercise contexts, the data are strictly captured manually in the form of written notes by observers or instructors. In virtual exercise contexts, data are mainly captured by the simulation environment. In other contexts, like the complex MWX environment, many sources of data exist but in a variety of formats. Instrumented range systems provide the position location information (PLI) for tracking movement of entities and some level of automated adjudication for force-on-force direct fire engagements and some simulated indirect fire. Many events and adjudications are tracked in notebooks or spreadsheets by exercise control (EXCON) and O/Cs. To address the variety and non-standardization of data sources, the Analysis Dashboard is being designed to auto-ingest from the range of data sources organically available at Marine training sites, depending on resource availability, for visualization and analysis. Also, standardized data documentation mechanisms have been created to capture and adjudicate significant events. The ingestible sources for unit location data can include any of the following:

- **Instrumented Tactical Engagement Simulation System (ITESS)**: When equipped with weapon-mounted laser emitters for force-on-force exercises, fire events and automated entity state adjudication are tracked in addition to position tracking. The system is radio-based and real-time tracking of entities is dependent on radio coverage. Data are stored and later downloaded during gaps in connectivity.
- **Tactical Assault Kit (TAK)**: TAK is an ecosystem for sharing tactical data. When used by exercise participants or O/Cs, it provides a means to capture entity PLI from mobile devices via commercial mobile or tactical networks connected to a TAK server.
- **GPS Tracking Devices**: GPS tracking devices have the advantage of portability, low cost, and tracking capacity when radio and mobile data coverage is poor and is limited by battery life and the inability to download data until the end of the exercise.
- **Track Log Data**: Manually-captured PLI as reported by periodic position reports.

Furthermore, a spreadsheet-based Significant Actions and Adjudication Tracker (SAAT) data format was developed to encompass significant events traditionally captured in various formats including notebooks, spreadsheets, or not at all. Initially demonstrated with fires events, it has evolved to include such events as intelligence, direct and indirect fires, logistics, electronic warfare, and other event types desired to be logged, tracked, and reported. Timestamps and location information are used to synchronize with entity tracking data for event visualization. Combined with starting strengths of units, critical items and equipment, and casualty adjudication, the event data is used to analyze the outcome measures of lethality and survivability.

Figure 2. Fires Event Fused with Entity Tracking Data



Fusion of the instrumented position tracking data and spreadsheet-based event logging and adjudication data from the SAAT provides a more complete picture of the events on the battlefield, both for exercise replay and for decision-making analysis. Figure 2 illustrates the combination of entity tracking data from ITESS (individual blue icons) with fires information ingested from a spreadsheet (red circle with associated metadata).

Cognitive Measures

Cognitive measures are quality ratings of individual or unit decisions and adaptations; for example, O/C ratings of the quality of the commander's guidance to the staff, exploitation of a situational change, or weapon to target match. Cognitive measures utilize five-point Behaviorally Anchored Rating Scale (BARS; e.g., Muchinsky, 2003) items where the behavioral anchors for a "1" represent novice performance and a "5" represent expert performance.

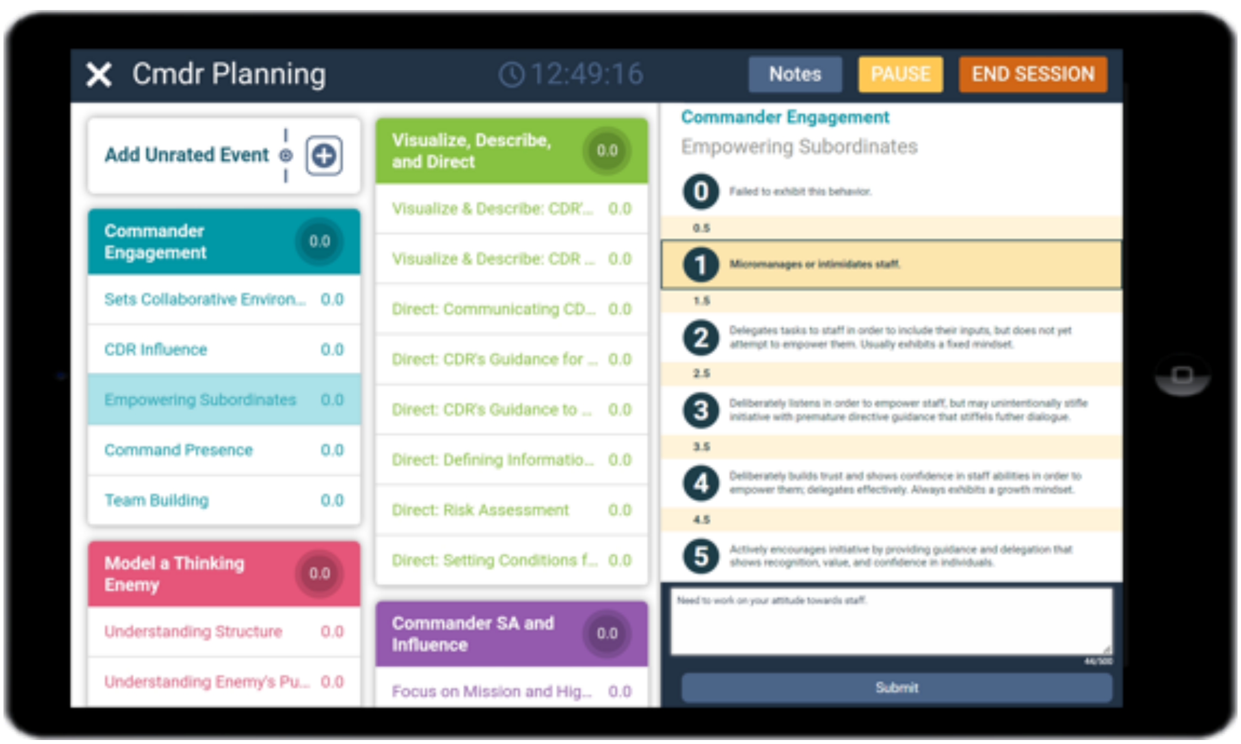
Table 1. Commander and Staff Behaviorally Anchored Rating Scales

Rubric	Key Performance Area	BARS Items
Commander	Visual and Describe the Battlespace	17
	Commander Engagement	5
	Model a Thinking Enemy	5
	Commander SA and Influence	11
	Commander's Role in Problem Framing	5
Staff	Tactical Competency	7
	Collaboration	8
	Functional Integration	8
	Organizational Effectiveness	3

The BARS items are grounded in the five-stage model of cognitive skill acquisition (Dreyfus & Dreyfus, 1986) which posits that distinct characteristics of performance are observable in the behavior of individuals at each of the five developmental stages reflecting different levels of knowledge and know-how. Thus, applying a BARS rubric to assess the quality of one's performance supports diagnosis of the individual's current stage of development.
Three BARS rubrics were developed for battalion and above (1) commander planning, (2) commander execution, and (3) staff processes. The BARS items span five Key Performance Areas (KPAs) for commanders and four KPAs for staffs (see Table 1). Initial BARS items for logistics personnel have also been produced under this effort. Previous research validated BARS for squad leaders and platoon commanders (e.g., Phillips et al., 2013) and all-source intelligence analysts (Phillips et al., 2021), which may be applied in future SMART-Viz exercises.

The BARS rubrics are authored onto the Field Assessment System prior to the exercise for employment by O/Cs in the field. Figure 3 depicts the Field Assessment System rating interface. Observers select the one behavioral anchor that best matches the observed performance and submit additional comments if desired. Performance ratings are made as performance is observed, in stride. That is, as opposed to generating a single rating per item of how a commander or staff performed across all five days of the exercise, items can be rated when they are observed, resulting in timestamped observations of decision-making performance. The timestamped ratings enable quality of decision-making to be aligned to specific exercise events, particular days of the battle, or any other desired timeframe to provide insight into the KPAs that contributed to effective or ineffective performance, or other factors that contributed to changes in KPA scores (such as sleep deprivation). When performance not captured by the BARS rubric is observed, O/Cs can create "unrated events" on which to comment. The Field Assessment System rating results are then exported as JSON or CSV files for analysis and manipulation in the Analysis Dashboard or other analytic tools.

Figure 3. Behaviorally Anchored Rating Scale Rubric in Field Assessment System



Outcome Measures

Outcome measures are measures of battle effectiveness. To date, *lethality* and *survivability* are being calculated and displayed. Lethality is reported for units down to company echelon (e.g., an individual battalion or battery) as well as for weapon systems (e.g., the F-35 or the High Mobility Artillery Rocket System [HIMARS]) so that participants can examine which assets achieved the most impacts. Users can choose to view lethality measures for the entire exercise or within a selected timeframe or geography, for example, during the engagement that took place on Day 2 at Site X.

Similarly, survivability is measured for personnel as well as for critical items, that is, certain assets deemed to be of sufficient value and therefore desirable to track such as 81mm mortars or Javelins. The outcome measure of *mission accomplishment* will be implemented in the next year of the project.

OODA Decision-Making Measures

Measures of adaptability originally validated in a controlled simulation (Daly et al., 2022) were modified for use in an OODA loop (see Hammond, 2018) visualization. The OODA measures include:

- **Observations:** Timestamps for and counts of the observations made by the different sensor types, such as electronic intelligence (ELINT), unmanned aerial sensor (UAS), fixed- or rotary-wing assets, a ground unit observer, or a counter-battery target identifier.
- **Decisions:** Timestamps for when an action is directed to be taken, such as issuing a target package, a call for fire, or battle damage assessment (BDA) tasking.
- **Actions:** Timestamps for when an action occurred as a result of a decision, such as the initiation of a fire or air mission, and the rounds complete for such missions.
- **Observe Time:** The duration from the first to the last observation of a particular adversary position, by the functional cell (e.g., intelligence or operations) overseeing the observations.
- **Orient Time:** The duration from the first observation of the adversary position to a decision about that position by the functional cell. It reflects the time spent by the cell to achieve situation awareness.
- **Action Time:** The duration from start to end of an action event, such as a fire, air, or maneuver mission.
- **Kill Chain Time:** The duration from the first observation to the end of the last action across the kill chain for the unit. It represents the unit's overall OODA timeframe.

To facilitate capture of only the data needed to produce an OODA kill chain, an OODA Observer Markings rubric was produced in the Field Assessment System. The Observer Markings rubric is organized by Observe, Decide, and Act inputs such that an O/C or a member of the AAR cell can efficiently input the variety of types of observations, decisions, and actions that occur with a timestamp and add their supporting comments. The comments specify the unit or asset associated with the event. To facilitate production of the OODA visualization, a Python programming language script was developed to automatically ingest the data output file from the Field Assessment System Observer Markings rubric and produce a visual representation of the OODA kill chain and the accompanying measures.

RESULTS IN TACTICAL TRAINING VENUES

The AAR tools and performance measures have been applied in live force-on-force exercises at both small-scale platoon echelon and large-scale regiment and division echelons. The purpose of this section is to describe how disparate clusters of data are merged into a comprehensive visualization of what occurred on the battlefield and how an analysis of unit performance is conducted.

Small Unit Field Exercises

The Basic School has utilized the data-driven AAR tools during seven evolutions of the War Field Exercise (FEX) to support planning, EXCON, and AAR for students and staff. War FEX is a 72-hour force-on-force field exercise that serves as the culminating event for Marine officer students, who are divided into six platoons. The War FEX incorporates the performance measures and visualizations to enhance the way students learn through the successes and failures of their decisions and tactics as platoon commanders and squad leaders.

The small-scale TBS setting relies primarily on instrumented unit location data from ITESS and the SAAT to visualize the behavioral data and measures. The system setup for TBS War FEX events requires minimal resources. A standard capability set for TBS exercise control consists of one computer loaded with the Analysis Dashboard and the SAAT spreadsheet, two TV monitors, a Vibe smartboard, and 10 TAK-enabled devices. With this small logistical footprint, the AAR tools can tie into any pre-existing or field expedient exercise control center.

Initially, TBS instructors used the visualization of unit movements and engagements in the Analysis Dashboard to conduct the AAR, moving to desired points on the exercise timeline to discuss key decision points. The procedure has now evolved into students taking on the exercise control and AAR responsibilities. Once the system setup is complete, exercise control students (i.e., those unable to participate in the field due to injury or illness) are taught how to utilize the Analysis Dashboard and fill out the SAAT, create additional visualization overlays on the map, conduct basic troubleshooting procedures, and load the SAAT into the Analysis Dashboard. Then, prior to execution, each platoon's plan is drawn as an overlay in the Analysis Dashboard using operational symbols and graphics. Drawing the plan enables comparison of planned and actual events during the AAR, as illustrated in Figure 4. Immediately following execution, students ingest the data sources and the Dashboard fuses data from ITESS and the SAAT.

The students create, rehearse, and conduct the AAR within 2-3 days of the exercise conclusion. Designated student leaders are given a one-hour class on how to conduct an AAR and are equipped with laptop computers to collaborate with their peers in the Analysis Dashboard, where they review the battle and use the behavioral data and measures to identify decision and learning points. Instructors guide and mentor students during a four-hour rehearsal on how to articulate their lessons learned to a large audience of approximately 300 student peers.

Figure 4. Analysis Dashboard with PLI Data and Operational Graphics Overlay

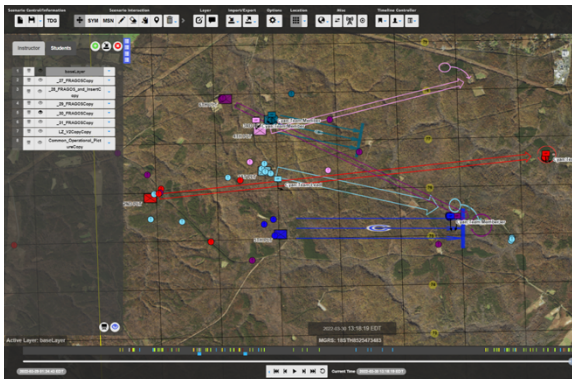
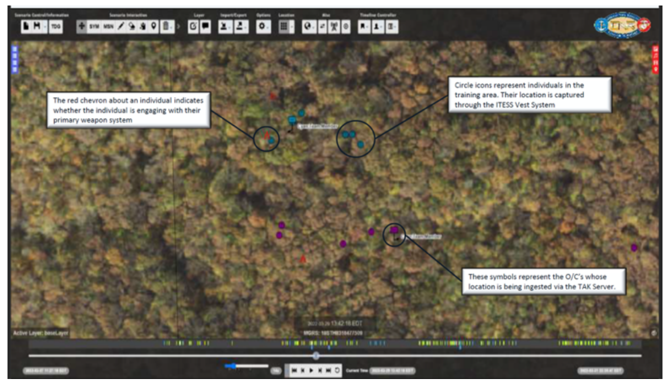


Figure 5. Analysis Dashboard Visualization of Platoon on Platoon Engagement



During one AAR, a student described how he chose to defend against the enemy's most likely course of action and was caught by surprise when the enemy executed a different maneuver. Utilizing the Analysis Dashboard visualization, the student replayed the event and noticed a small gap in his formation which he failed to cover with fire or observation. Another student from the same class identified his decision point to conduct a spoiling attack after replaying the battle footage. In both examples, students identified their own successes and failures, helping them improve their tactical cunning for future training and combat operations. Figure 5 illustrates one War FEX engagement and highlights the visualization of the two engaged platoons.

Large Scale Field Exercises

The performance measures have been identified, produced, and piloted in the MWX environment in collaboration with the AAR cell. The AAR cell goals are to use data to identify the decisions, contributing factors, and outcomes associated with key events, and to illustrate key events to the exercise participants in the AAR to promote learning.

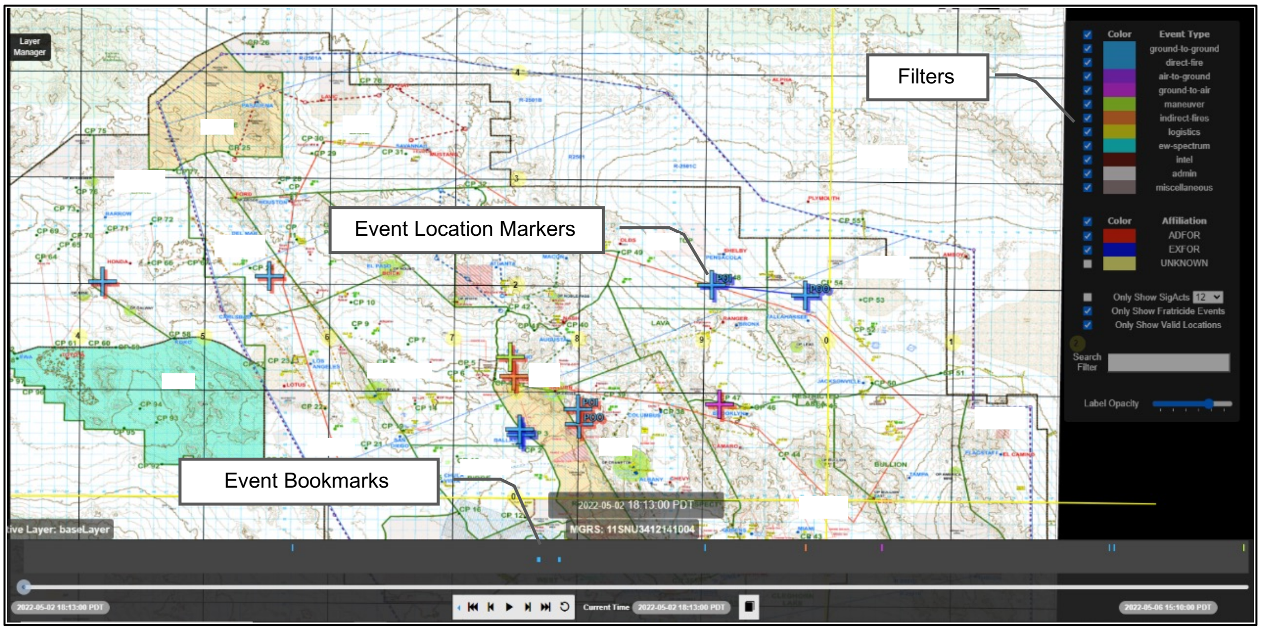
The MWX is the culminating event of the Marine Corps Service Level Training Exercise. It is a five-day free play force-on-force exercise that varies in scope and mission depending on unit availability. Ground force participation can include up to a Marine Division with three Infantry Regiments, an Artillery Regiment, and a Combat Logistics Regiment. The exercise also includes various elements of a Marine Corps Air Wing and accompanying command and control (C2) elements. Many sorties may be flown in the five-day period in support of the ground forces. Given the enormity and free-play nature of the exercise, the challenges associated with capturing objective ground truth and standardizing performance measurement across exercise evolutions are readily apparent.

Behavioral Measures, Event Tracking, and Adjudication

Significant event data evolved over several exercises, with the initial proof of concept used to visualize the time and location of each indirect fire event captured in a spreadsheet to supplement the instrumented range data already being visualized in the ITDG application. Building on the success of the fires visualization, the AAR Working Group developed the SAAT spreadsheet format to capture all significant events to be tracked at the MWX. The result of using this data capture, as shown in Figure 6 showing fratricide events, allows the visualization in time (bookmarks) and space (event markers) of all significant events meeting criteria applied through filters. The user can utilize the markers, bookmarks, and event details to zoom in on the event and inspect unit movements and other activity in the vicinity of the time and location of the event itself, facilitating analysis and AAR discussion.

Additionally, multiple events can be chained together into a story to link events related by a common mission, region, event type or other criteria and automatically play out the events in series, facilitating presentation of the entire chain of events. This feature aids in not only facilitating AAR presentations, but also building take-away packages for the units to review and analyze at their home stations.

Figure 6. SAAT Event Visualization



Finally, to better aid analysis of lethality and survivability, more detailed adjudication data previously only available in descriptions was added to the significant events, with Tables of Organization and Equipment (TO&Es) for Blue and Red units. TO&Es detail starting quantities of personnel and equipment per unit and critical items to be tracked. With initial experimentation during small unit training at TBS's War FEXs proving successful, the fully-evolved SAAT was used at the most recent MWX to track a large exercise with over 600 tracked events.

Cognitive Measures for Assessing Decision Making

The three BARS rubrics were applied at one MWX to pilot test the approach for assessing the decision-making performance of battalion and regimental commanders (N = 3) and staffs (N = 23) during planning and execution. A total of 181 commander and 778 staff ratings were collected over the course of one exercise to measure performance on the KPAs. Data files were exported from the Field Assessment System and imported into a Power BI dashboard to visualize the data and provide descriptive statistics.

The BARS items have not yet been validated due to the small sample size, thus the outcomes reported here are provided to illustrate analyses that can be conducted post-validation. As shown in Figure 7, commanders were rated highest on the *Commander Situation Awareness and Influence* KPA and lowest on the *Model a Thinking Enemy* KPA. Notably, performance seemed to improve from planning to execution, a finding opposite to expectation since execution is the more dynamic of the environments.

Similarly, staffs were also rated higher during execution than planning. More detailed analysis of individual staffs indicates variability in KPA scores (Figure 8, with EXCON units in blue and ADFOR unit in red) and illustrates the forthcoming capability to provide data-driven feedback to units based on their relative strengths and deficiencies.

Figure 7. Average KPA Ratings of Commanders in Planning and Execution

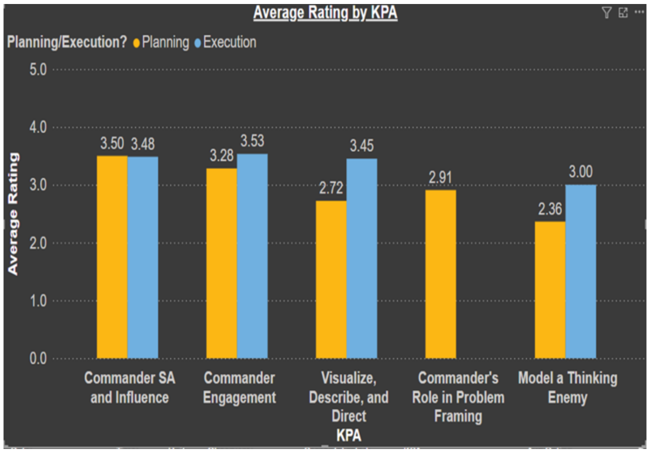
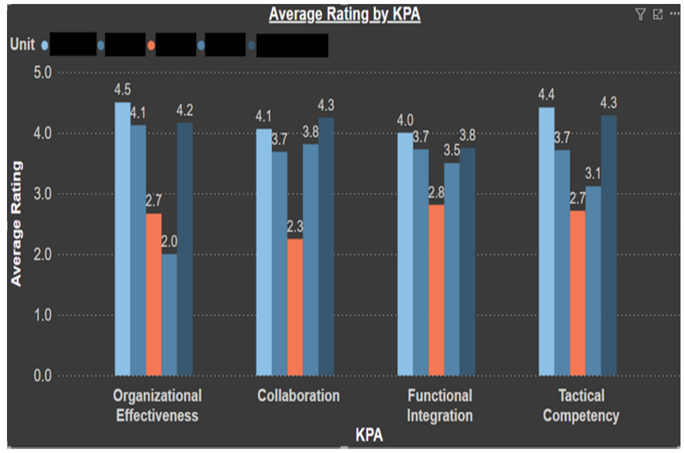


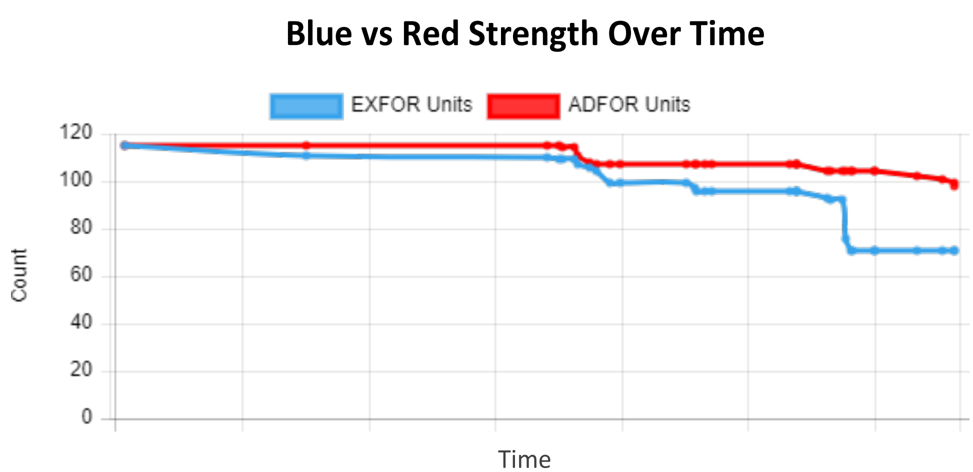
Figure 8. Average KPA Ratings of Regimental Staffs During Execution (redacted)



Outcome Measures

Lethality and survivability outcome measures were produced at one MWX. The lethality and survivability measures are calculated based on SAAT data. Figure 9 illustrates the representation of survivability over time for Blue versus Red forces.

Figure 9. Survivability Over Time



OODA Decision Making Visualization and Measures

The first OODA fires kill chain was produced manually at one MWX to detail the destruction of a mobile forward C2 node. Timestamps for the Observe, Decide, and Act events within the C2 node kill chain were collected and a visualization was produced to show the series of events contributing to the kill chain. Both units benefited from the visualization of events. For Unit B who executed the kill chain, it was informative to view the multiple OODAs of the elements involved in the chain of events: the intelligence function which senses and confirms the enemy position; the operations function which decides what to target and with what weapon system; and the battery that prosecutes the target. The time it took the elements to conduct each step of their respective OODAs supported detection of communication delays, process obstacles, and uncertainty tolerance. It was also informative to see the number of sensors involved to pinpoint the enemy position prior to attack. For Unit A that was sensed and destroyed, the greatest insight was the time window during which the node could have been repositioned to avoid destruction.

The OODA visualization and time-based measures resulted in learning by the unit commanders and adjustments to future behaviors. During the following MWX six months later, commanders exhibited two new practices to increase their C2 node survivability. As a force protection measure, they collected geospatial intelligence and ELINT *on their own positions* to monitor what the adversary could sense about them. When they were observable by those sensors, they moved. Notably, no C2 nodes were either targeted or destroyed in that MWX. Further, as a practice to speed their OODA cycle, the commanders minimized the number of observations deemed sufficient to prosecute a target.

CONCLUSIONS AND FUTURE DIRECTIONS

Efforts to date provide anecdotal evidence that data-driven AARs supported by visualization and analytics enhance learning and understanding of tactical decision making. In the small-scale exercise setting at TBS, the visualizations of the battlefield enable deep understanding of tactical principles, dynamic decision-making, assuming risk, sensor capabilities, and detection avoidance. In the large-scale MWX environment, the battlefield visualizations provide awareness of ground truth and enable AAR facilitators to drill down on events and build vignettes to illustrate how decisions and actions produced consequences over time. Furthermore, introduction of the OODA representation of fires kill chains has resulted in effective decision-making and process adjustments by commanders.

The scalability of the performance measures and AAR tools across exercise environments has been a primary driver of measurement and development approaches. To support performance measurement across echelons, scenarios, and exercise settings, a broad range of behavioral, cognitive, outcome, and OODA decision process data and measures have been identified and captured by the AAR tools. This range of data availability supports an AAR facilitator to both identify key events for the AAR and select the visualization of the battlefield, analytics, or decision process most illustrative of the learning point. Furthermore, every exercise setting, from TBS to MWX and live to virtual, differs in its resources and exercise control process, and thus the data sources available to produce the visualization and analytics. To provide the performance data and measures that drive the AAR, the Analysis Dashboard is designed to ingest a range of data sources to obtain the necessary inputs to the visualization and analytics, from instrumented systems providing PLI or electronic spectrum emissions, to basic Excel event logs.

Introduction of the cognitive measures to assess decision-making with BARS is under development yet gaining significant support from MWX and MCTOG personnel. Use of a single BARS rubric across multiple O/Cs provides a standardized approach to assessment. This allows commanders and staffs to receive feedback on a common set of agreed upon performance areas vice rich but potentially skewed feedback stemming from individual O/C biases. Furthermore, a standardized set of cognitive measures, once validated, can be applied more broadly across formal school and exercise settings to track individual progress, performance trends, and training effectiveness.

Future Directions

To further demonstrate the scalability of data-driven AAR tool visualization and analytics supporting learning, the next step is to conduct experiments at medium level live exercises such as a Marine Corps Combat Readiness Evaluation (MCCRE). The MCCRE is a required event for all Battalion level commands to evaluate unit combat readiness before they deploy. Following the MCCRE test context, the feasibility and utility of the performance measures and AAR tools will be examined in virtual and constructive environments to further test the measures and tools at different exercise sizes, locations, and constructs across the Marine Corps.

The data-driven visualizations and analytics can be applied for numerous purposes beyond the learning of exercise participants during the AAR. For example, a take-home package of tactical visualizations and analytics showcasing why tactical maneuvers were met with success or failure could be referenced infinitely by the exercising unit but could also be used to prepare incoming units to future MWXs. The commander could apply another unit's lessons learned to improve their plan, and the Marine Corps as a force could continue to raise the performance bar rather than repeating the same mistakes. This use case was demonstrated by the MWX commanders who adjusted their force protection and targeting approaches based on the OODA visualization of a previous unit.

Further, the utility of the exercise performance data can be extended to institutional questions, such as course effectiveness. Tracking the performance of exercise participants over time reveals common misconceptions and decision-making challenges that schools can address in their curricula. MCTOG is investigating the usage of common cognitive measures in courses and MWXs for the purpose of assessing transfer of course lessons to field environments.

Beyond performance assessment in a singular exercise environment, the analytics offer an opportunity for the Marine Corps to apply consistent cognitive and behavioral measures across an individual's career or a unit's training life cycle. Tracking performance and learning data over time supports identification of Force-wide trends in learning effectiveness. In conjunction with the Commandant of the Marine Corps' guidance to focus on the use of machine learning, artificial intelligence, and predictive analytics to enhance learning potential, the performance data could be used by units or schoolhouses to analyze trends and shape future activities.

At the highest level, the ability to collect standardized, objective performance data supports Service-level experimentation with new processes, technologies, and materiel solutions. The effectiveness of tactics, training, procedures, C2 systems, force structures, and even human-machine teaming can be tested by comparing unit combat performance and decision quality with their usage to the baseline capability. Furthermore, given sufficiently large data sets, artificial intelligence could be employed to identify non-obvious associations among behaviors, decisions, and battle outcomes, as well the interplay among the decisions and actions-reactions-counteractions of opposing forces.

The lessons learned for instituting a data-driven AAR that optimizes insight and learning can be applied to other AAR settings. The approach to measuring decision-making performance and battle outcomes is universally applicable. The process for enabling visualization of key battlefield events supported by event data, regardless of data source availability, can be extended to other tools. Finally, the effective combination of instrumented and human data inputs providing performance analytics bears replication in other similar AAR settings.

ACKNOWLEDGEMENTS

This work was supported by Dr. Peter Squire and the Office of Naval Research. The authors thank Dr. Squire and his team; LtCol Jesse Attig and Mr. Ryan Brown; and the many Marine Corps Air Ground Combat Center contributors.

The views of the authors expressed herein do not necessarily represent those of the U.S. Marine Corps, U.S. Navy, or Department of Defense (DoD). Presentation of this material does not constitute or imply its endorsement, recommendation, or favoring by the DoD.

REFERENCES

Campbell, G., Zlatkin, A., & Riley, J. (2021). Interactive military instructor training and assessment technology (IMITATE): Phase II Final Report (Report under contract N00014-17-C-1031 for the Office of Naval Research). Design Interactive, Inc.

Daly, T., Hancock, A., Phillips, J. K., & Mainz, M. (2022). Measuring small unit adaptability in a virtual environment. Proceedings of the Interservice/Industry Training, Simulation, and Education Conference. NTSA.

Department of the Army (1993). A leader's guide to after-action reviews (TC 25-20). https://www.acq.osd.mil/dpap/ccap/cc/jcchb/Files/Topical/After_Action_Report/resources/tc25-20.pdf

Dreyfus, S. E. & Dreyfus, H. L. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* The Free Press.

Hammond, G. T. (2018). A discourse on winning and losing. Air University Press.

Kaliner, J. (2013). When Will We Ever Learn? The After Action Review, Lessons Learned and the Next Steps in Training and Educating the Homeland Security Enterprise for the 21st Century. Thesis, Naval Postgraduate School, Monterey, CA.

Muchinsky, P. M. (2003). *Psychology applied to work*. Melmont, CA: Wadsworth/Thomson Learning.

Phillips, J. K., Hancock, A. K., Ross, W. A., Williams, M. L., Daigle, R. N., Pfahler, M. D., Baxter, H. C., & McGee, R. G. (2021). A mastery model to support Marine Corps ISR enterprise workforce professionalization. (Report under contract N00014-18-C-2034 for the Office of Naval Research). Cognitive Performance Group.

Phillips, J. K., Ross, K. G., Rivera, I. D., & Knarr, K. A. (2013). Squad leader mastery: A model underlying cognitive readiness interventions. Proceedings of the Interservice/Industry Training, Simulation, and Education Conference. NTSA.

Young, C., Schaffer, R., Longtin, M., Stensrud, B., & Mainz, M. (2019). Tactical Decision Kits for Infantry Training. Proceedings of the Interservice/Industry Training, Simulation, and Education Conference. NTSA.