Providing Better Feedback to Aviators through Automated Human Performance Analysis

Robert Siegfried, Torsten Müller
Aditerna
Riemerling/Munich, Germany
robert.siegfried@aditerna.de, torsten.mueller@aditerna.de

Mark Burgess
Prevailance Aerospace
Virginia Beach, VA
mark.burgess@prevailanceaerospace.com

Krzysztof Rechowicz
Old Dominion University
Suffolk, VA
krechowi@odu.edu

ABSTRACT

The evaluation of trainees and aircrews requires highly trained subject matter experts (SMEs) to assess current performance and recommend future training that is optimized for each individual. An increasing volume of data generated from today's training environments makes a comprehensive and consistent trainee assessment impractical for the instructors, if not impossible. In addition, training facilities are faced with a shortage of qualified SMEs and instructors, which reduces time for properly assessing individual performance even more. We fight as we train, so unsatisfactory trainee assessment and missing individual training recommendations inevitably yield warfighters that are not proficient enough for the 21st century military operating environment.

To relieve instructors from routine tasks and to support training facilities, we have investigated the use of automated human performance analysis and individualized training recommendations. For this, we have designed the Fleet Operational eXercise Training for Warfighter Optimization (FOX TWO) prototype that ingests training data, calculates and visualizes predetermined Measures of Performance, and then provides individually adjusted training recommendations. The automated computation of performance metrics allows instructor to provide more immediate and consistent performance assessment, while freeing up time to focus on more advanced evaluation and feedback. Included is a training data store that allows training facilities and instructors to track trainees throughout their career, and to identify when a skill is mastered and how often trainees need to practice that skill to stay proficient.

This paper presents results from working with Naval Aviation, including concept design for adaptive data storage and retrieval, flexible data analysis mechanisms and validation through prototype implementation and human subject experimentation. Human subject experimentation to verify the underlying concept to make individual training recommendations included 20 aircrew trained, 100+ simulator events executed with 700+ scenario executions that resulted in 7+ GB of training data collected.

ABOUT THE AUTHORS

Dr. Robert Siegfried is a Senior M&S Consultant and Managing Director of Aditerna. Robert is serving as Chair of the NATO Modelling and Simulation Group (NMSG) and leads the NMSG's efforts towards Modelling and Simulation as a Service (MSaaS). Robert is a member of the Executive Committee of the Simulation Interoperability Standards Organization (SISO) and is actively engaged in multiple SISO working groups.

Torsten Müller is Managing Director of Aditerna and certified project manager with 15+ years of experience in various international projects and is a recognized expert for IT project management in traditional and agile project environments. He has successfully led various projects in different domains, such as Telecommunications, Transport, and Defense.

Mark Burgess Director of Operations and Chief Instructor Pilot for Prevailance Aerospace. Mark served 20 years as a Naval Aviator and deployed for three combat cruises. Prior to Prevailance, he provided contract support to OPNAV N98 in the Aviation Warfare Training Division.

Dr. Krzysztof Rechowicz is a research assistant professor at the Virginia Modeling, Analysis and Simulation Center (VMASC) at Old Dominion University (ODU). His research interests focus on making the world more inclusive through sensory and reality augmentation. He is currently the Digital Senses Lab lead. Krzysztof is also an IEEE Senior Member and a member of the Society for Modeling & Simulation International.

Providing Better Feedback to Aviators through Automated Human Performance Analysis

Robert Siegfried, Torsten Müller
Aditerna
Riemerling/Munich, Germany
robert.siegfried@aditerna.de, torsten.mueller@aditerna.de

Mark Burgess
Prevailance Aerospace
Virginia Beach, VA
mark.burgess@prevailanceaerospace.com

Krzysztof Rechowicz
Old Dominion University
Suffolk, VA
krechowi@odu.edu

1. PROBLEM STATEMENT

1.1 Need For Change

Emerging warfare capabilities offer many novel tactical options to commanders across all facets of combat (Cohen et al., 2020; Sayler, 2020). However, the dynamic and complex nature of integrating these new capabilities into existing operations results in a multitude of training challenges (Freeman & Zachary, 2018). Realistic simulated training becomes paramount as the complexity of tactics, techniques, and procedures (TTPs) increases. Along with the associated increase in simulated training comes the challenge of building and choosing the best scenarios for the training objectives.

Thorough trainee evaluation requires subject matter experts (SMEs) to assess current performance and recommend future training for the individual. With limited SME availability and large volumes of data generated from today's training environments, SMEs' thorough and consistent trainee assessment is impossible (Fan, Han, & Liu, 2014; Hodson & Hill, 2014; Kitchin, 2014; Labrinidis & Jagadish, 2012; Song, Wu, Ma, Cui, & Gong, 2015). This leads to the common situation where very little analysis is done to determine a trainee's proficiency, select the most appropriate future training, and evaluate if a scenario truly ensures training objectives are mastered. Without this knowledge, choosing a follow-on training scenario tailored to that individual and ensures mastery of previous deficiencies is impossible.

Working with the Naval Air Warfare Center Training Systems Division (NAWC TSD), the objective was to design and develop a software technology that leverages data science, artificial intelligence, and advanced computational analyses of tactical data sources to improve training assessments and to automatically select future scenario recommendations that make training more adaptive, efficient, and effective.

1.2 Naval Aviation

The authors looked at Naval Aviation's current problems and how the investigated approach can help answer those questions. The remainder of this paper builds upon experiences and lessons learned in the Naval Aviation community. Although some observations may be specific to Naval Aviation, most results presented in this paper can also be transferred to other communities.

Previously, Commander Naval Air Forces (CNAF) stated that two of his most challenging issues to solve were generating current readiness and recovering readiness after a post-deployment stand-down. On March 21, 2018, the current CNAF stated, "The mission of the Naval Aviation Enterprise is to sustain required current readiness and advance future warfighting capabilities at best possible cost." (Miller, 2018) To complicate this task further, upgrades in aircraft weapon systems require more networked and integrated tactics and training.

To answer these problems, Naval Aviation developed the Naval Aviation Simulator Master Plan (NASMP). NASMP's goals included upgrading simulators to allow for Training and Readiness (T&R) to be achieved in them, integrating platform simulators to each other and other platforms, and integrating simulators with live ranges and aircraft. These advances will greatly enhance Naval Aviation's ability to achieve readiness faster and cheaper than ever before. Unfortunately, these changes will lead to enormous requirement for SMEs to build new simulator training scenarios for integrated training. SMEs will also be responsible for analyzing and modifying the scenarios based on their expert analysis. This will be in addition to their current responsibility of analyzing how well their trainees are performing. Ultimately, this will lead to a significant increase in SME demand.

A second issue that the authors identified is the need for a new system to evaluate how T&R is achieved and funded in Naval Aviation. Both flight time and simulator hours are funded based on an antiquated T&R system. Since the implementation of NASMP, very little study was performed to evaluate the differences in how trainees practice and learn in a simulator versus an aircraft. Also, the current T&R system only assesses a trainee's currency in a skill and not their level of proficiency. Only by assessing how proficient a trainee is at a skill (whether in the air or simulator) can SMEs determine how often they need to retrain.

1.3 Challenges of Today's Training Cycle

In today's live and simulated training, Measures of Performance (MOPs) are used to analyze how well an individual or team performed, document the training, and track their training progress and current readiness. Unfortunately, little analysis is done to determine if the chosen training scenario ensures the trainee masters the tactic/procedure or if there is a more effective scenario for that training objective. Figure 1 conceptualizes today's typical process of training delivery and subsequent scenario improvement.

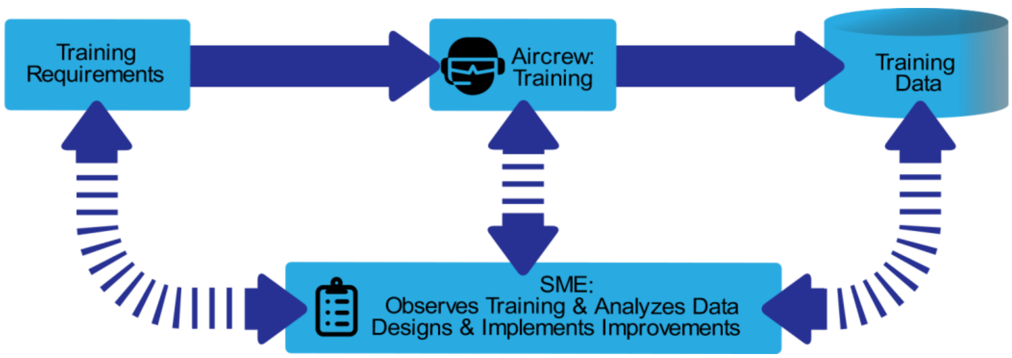

Figure 1. Typical training adaptation and improvement process today (simplified).

The process begins with a student who needs to train to meet a specific task, Training & Readiness (T&R) requirement, Air Combat Training Curriculum (ACTC), etc. That student's training is typically led by an SME (e.g., Training Officer, Simulator Instructor) who must develop a training scenario applicable to the requirement. That SME must also develop the necessary training products for all ancillary simulated participants to add realism to the training scenario. Products that add to the scenario's operational realism include Special Instructions (SPINS), Air Tasking Order (ATO), game plans, timing, Red Air game plan, and many more. Simulator training also requires initial conditions and parameters for the Semi-Automated Forces (SAF) or Computer-Generated Forces (CGF) to be effective.

Once training takes place, a large quantity of data is generated that is rarely saved longer than to be used for the event debrief and rarely analyzed against other training variations. In general, the only products saved are grade sheets and hand-written lessons learned. Unfortunately, neither the grade sheets nor lessons learned are typically used outside the unit being trained. Any changes to future training are made by the SME and are limited by the SME's available time and focus. The SMEs must use their judgment to determine how well the scenarios served to train to the required task compared with other possible scenarios.

Historically, scenarios are designed to train to the task required, with little comparisons made to variations in the scenario that might make the training more effective, efficient or adaptive. As long as the training requirements are met, the status quo is the easiest way forward. Redesigning the scenario requires hours of work by an SME and acts as a deterrent.

The current training adaptation and improvement cycle has multiple shortfalls and is insufficient to deliver the amount and quality of training that is required to meet the demands in a rapidly changing security environment. The two main challenges, in our opinion, are:

Challenge 1 – Dependence on Manual Tasks (Insufficient Automation): The majority of activities to prepare, execute and analyze a training event for a specific student, as well as the overarching activities to review and improve training scenarios, are manual tasks. Little, if at all, automation is currently brought to bear to relieve SMEs from recurring or dull tasks. An average Navy helicopter squadron has 12 pilots, each training 3-4 events per week, which sums up to 40-50 training events per week for a single squadron. With usually 1-2 Training Officers per squadron, each officer has 20-25 events per week to review. Obviously, and even without taking holidays etc. into account, this demonstrates the need to relieve SMEs from any 'automatable' task.

Challenge 2 – Too Few Subject Matter Experts (SMEs): The entire training cycle is dependent on the availability of highly skilled SMEs. Suitably qualified personnel ideally combine expertise in operational requirements and procedures, instructional design and simulator technicalities (e.g., how specific effects can be represented in the training environment).

2. SOLUTION APPROACH

In a recent research effort for the U.S. Naval Air Warfare Center Training Systems Division (NAWC TSD) Team Prevailance, composed of Prevailance, Aditerna and Old Dominion University's Virginia Modeling, Analysis and Simulation Center (VMASC), investigated how the outlined training challenges may be addressed.

2.1 General Approach

Figure 2 illustrates the basic approach to address the outlined challenges. The key idea is to introduce an automated data analytics solution into the training cycle that supports the SMEs through trainee performance analysis and individualized training recommendations. Essentially, the key objective is to free up valuable SME time to make sure they can properly focus on those tasks that cannot be automated but require expert judgment.

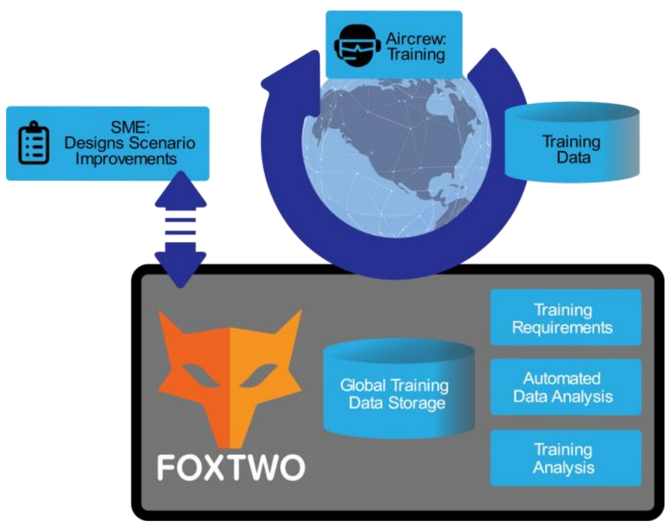

Figure 2. Data analytics as an integral part of an iterative training cycle.

The solution approach is referred to as Fleet Operational eXercise Training for Warfighter Optimization (FOX TWO). In a nutshell, the vision is that training data is automatically uploaded into the system, preprocessed (cleaned etc.) and then post-processed by various analyzers. These analyzers determine how well a trainee met predetermined MOPs. Through individual training performance tracking, follow-on scenario and training recommendations are derived that ensure trainees will master the required warfighting skills. Tracking trainees throughout their careers makes it possible to identify when a skill is mastered and how often trainees need to practice it to stay proficient.

2.2 Target User Communities

Although the authors primarily investigated the solution approach to assist training instructors and SMEs, further user communities might benefit as well, such as:

- The individual trainee can use such a system in a "self-service mode" to assist them in selecting the ideal training scenario to meet the training objective they are trying to achieve.
- Training and operations officers can use such a system to plan training events to provide the most efficient use of the available assets.
- Training centers can tailor training events to meet all users' objectives, maximizing available instructional time and minimizing product development time.
- Instructors can use such a system to vary a multitude of training parameters to provide scenarios that produce the most effective learning.
- Unit Commanders can use such a system to build training events that best integrate multiple assets under their command.

2.3 Key Components

Figure 3 illustrates the overall concept design of the solution approach. Essentially, the solution approach follows a three-tiered approach of data storage, business logic and user interface. To satisfy the unique needs, each of the three tiers has been specifically designed to match the requirements and provide optimal performance.

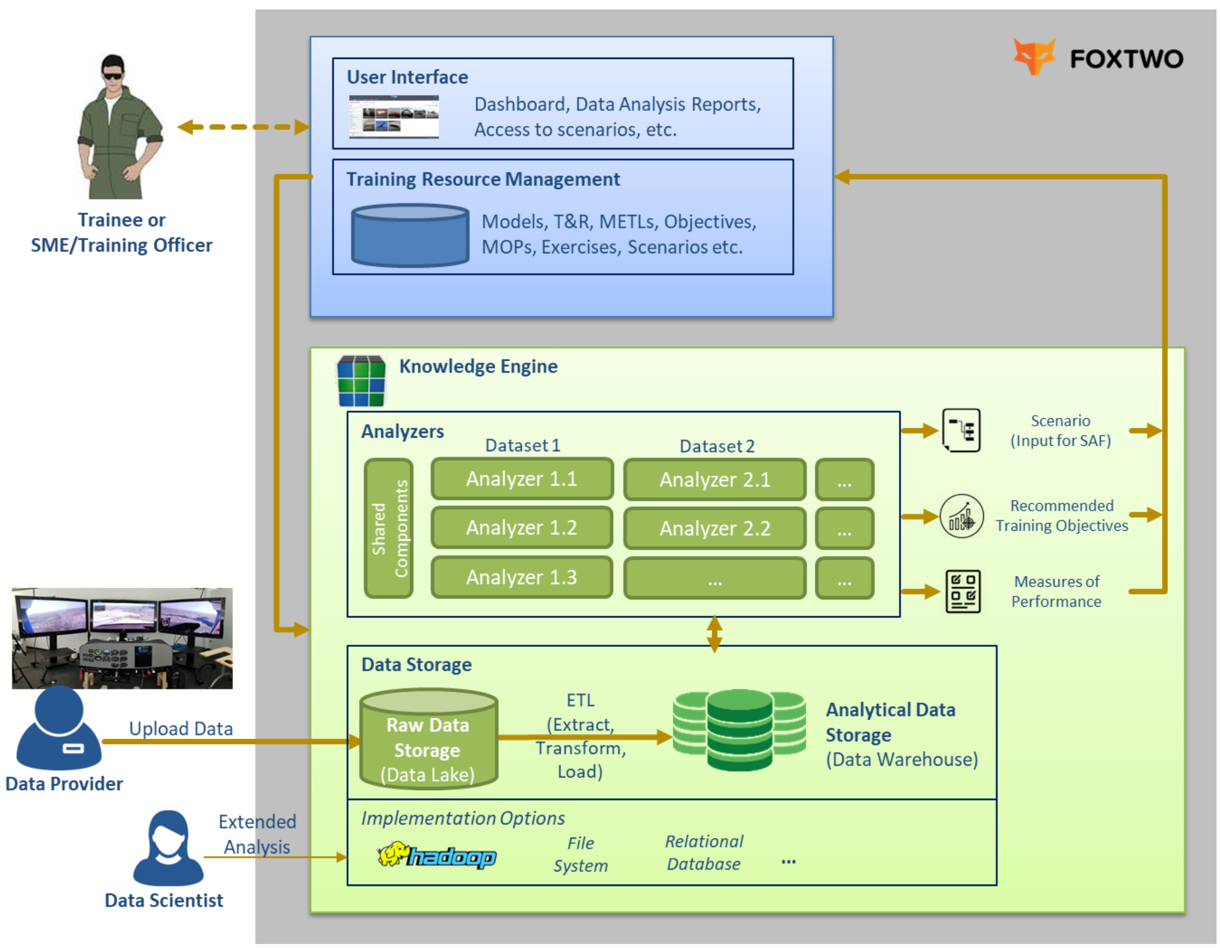

Figure 3. Solution Approach – Overall Concept Design.

2.3.1 Data Storage

Training datasets may either be manually uploaded by the user or directly transmitted from simulators and training devices through automated interfaces and data exchange mechanisms. All datasets are then stored in the Data Lake. The Data Lake acts as raw data storage, i.e., uploaded datasets are stored "as they are" in their original format. The data lake is introduced for two main reasons:

1. To keep the original data for later analysis. For example, new analysis techniques may be added at any time. As it is unknown in advance which portions of the original data are required (for yet unimplemented analysis methods), the only possibility to allow an analysis of previously uploaded data is to keep the raw data. To avoid problems due to preprocessing, transformation or filtering, the data lake stores the original data in its raw format.
2. Keeping the original raw data is the only way to enable backward traceability, i.e., the ability to provide detailed, explainable and understandable reasons how a specific recommendation or result was achieved.

Datasets are kept in the data lake for a specified amount of time. As old/outdated datasets may no longer be of value, those can be removed from the data lake. Actual data archiving or deletion policies will be defined together with the user.

Once datasets have been uploaded to the data lake, they go through the "Extract, Transform, Load" process and are transferred into the analytical data storage. The key idea of the analytical data storage is to preprocess the raw datasets from the data lake and transform them into a representation that allows appropriate analysis. The process includes the following steps:

- Data Validation - Datasets are validated if they comply with the expected format (folder structure, files, file types, etc.).
- Data Cleaning - Datasets are cleaned and filtered (i.e., only data required for subsequent analysis are moved to the next steps). Duplicate datasets are filtered out entirely during this step.
- Data Transformation - If required, data transformations are applied.
- Data Aggregation - If required, data aggregations are performed based on raw datasets.
- Data Loading - Finally, the resulting data is loaded into the analytical data storage.

The analytical data storage itself is implemented using different technologies. The reason here is that specific analysis techniques require data to be provided in a certain way. Some analysis techniques may work best if data is held in a data warehouse (e.g., built upon a relational database system). In contrast, others prefer data provided as a stream (e.g., from a Hadoop cluster file system).

2.3.2 Analyzers

Each dataset requires individual analysis methods and tailored algorithms. Therefore, the solution approach must support the use of multiple artificial intelligence-based (AI) techniques and data analysis approaches for each dataset, the so-called "Analyzers." Analyzers are (in general) specific for an individual dataset as they require certain domain knowledge. Common aspects that are reused in multiple analyzers are part of the "Shared Components" library.

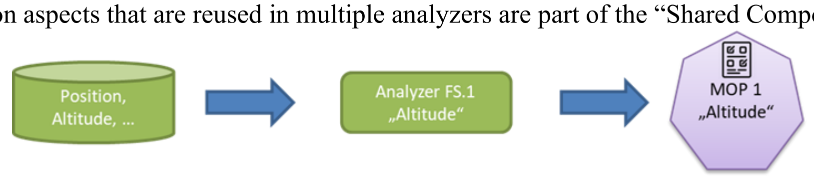

Figure 4: Illustration of Analyzer

Essentially, as illustrated in Figure 4, analyzers are self-contained program elements that take an input (i.e., analytical data), execute an algorithm and generate an output (e.g., a MOP). The key element here is that each analyzer may use the analysis technique or AI method best suited for the individual purpose.

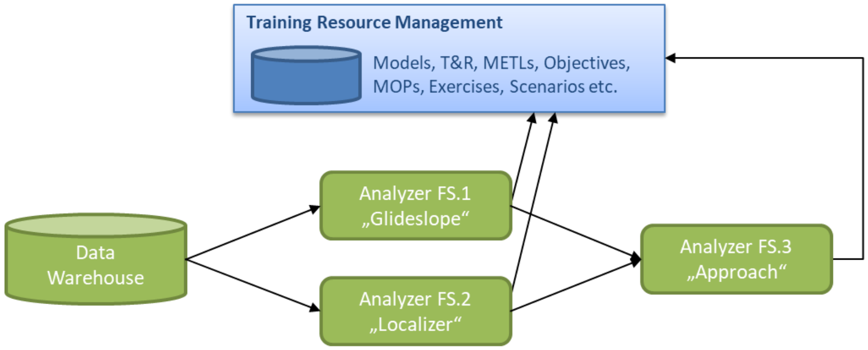

Figure 5: Simplified Example of Cascading Analyzers

Figure 7 illustrates cascading analyzers: the results of each analyzer may be used as input for further analyzers. In the example shown in Figure 5, the analyzer FS.3 requires two inputs. Each result (in this example, three MOPs are calculated for each training event) is stored individually and utilized separately to determine future recommendations.

Once new training data is uploaded and available in the analytical data storage, the appropriate analyzers are selected to generate output. The analyzers produce the following:

- Measures of Performance
- Scenario recommendations
- Training objectives

The results of the analyzers are fed back into the Training Resource Management module and are then available to the user. The analyzers themselves are version-controlled to support changing requirements (e.g., updated MOP definitions, T&R requirements). Modifying an analyzer thus results in a new version of this analyzer without breaking backward compatibility and traceability of MOPs.

As each analyzers is specifically designed to operate on a specific set of input data and to generate a specific MOP, the authors found it difficult to come up with general recommendations how to design and develop such analyzers. Best results were achieved by following a crawl-walk-run approach, that in this context translates to starting with rather simple analyzers (usually deterministic, e.g., using straight-forward time-series analysis) to get familiar with the actual analysis required for the current input data and desired MOP. Based on initial results, the analyzers can then be refined and extended. Initial lessons learned confirmed that more sophisticated (and more complex in term of development and testing) approaches, e.g., involving Machine Learning, are often not required.

2.3.3 Training Resource Management & User Interface

The Training Resource Management component allows the management of all training and exercise-related resources and products. This includes, for example (manual) definition of scenarios, upload of scenario products, and management of master data such as training objectives, Training and Readiness requirements, Mission Essential Task Lists (METLs), Measures of Performance (MOPs), Aircrew information (units, crew members). Users can add, modify and delete the information as required. Data can also be searched and accessed as necessary.

2.4 Prototype Implementation

A prototype implementation was developed to validate key aspects of the concept design, demonstrate technical feasibility, and evaluate operational value. As much as possible, existing software components have been reused to minimize implementation risks and shorten development time.

Specifically, Aditerna SRP was used to provide the Training Resource Management capabilities and the user interface. Aditerna SRP is a customizable commercial-off-the-shelf (COTS) product for training information management. A Hadoop Cluster includes data storage for raw data storage and a relational database management system (PostgreSql, in this case) for analytical data storage. The analyzers have been implemented in Java as an extension to the underlying SRP platform.

4. CONCEPT VALIDATION

4.1 Early-Stage Concept Validation

Concept validation in the early design activities was done with an interim prototype implementation using an example dataset from a real-time strategy game. Specifically, we have used a dataset including close to 8,000 matches from professional gamers leagues and international tournaments of “StarCraft Broodwar” (Synnaeve, 2018). Real-time strategy games involve human and robotic players, resembling virtual simulation training systems. The benefits of using game data are its public availability, without limitations or restrictions, and a large amount.

The early-stage concept validation focused on executing an end-to-end test of the envisioned solution approach, including all key components, such as the data stores, analyzers, training resources. The initial prototype was able to load approximately 15,000 game-related data files, which provided input for a set of analyzers that determine specific game strategies (decision tree-based). The early-stage concept validation demonstrated the feasibility of the solution approach and the overall concept design. No significant concept or design flaws have been identified.

4.2 Continued Concept Validation

A Cessna 172SP (C172) flight simulator was designed and built by researchers at VMASC and housed in the Digital Senses Lab for continued concept validation. The simulator was equipped with a standard physical cockpit configuration and force-feedback yoke. The simulator was driven by X-Plane 11 Global (www.x-plane.org), a retail version of the FAA-certified professional version. All parameters were baselined and tested to operate correctly. Three large video monitors, one in front and one on each side, were aligned to provide 180-degrees of visuals surrounding the cockpit setup. The avionics installed were upgraded and tested to ensure proper replication of existing aircraft systems. Local Cessna instructors and SMEs were brought in to validate the replication of the actual aircraft’s systems.

Figure 6. VMASC Cessna 172 simulator.

The C172 simulator gathered hundreds of parameters associated with the flight, including position, orientation, airspeed, glideslope, and localizer information. The dataset collected closely resembles data as operational flight simulators typically produce it. The simulator was also outfitted with a suite of sensors collecting data related to the pilot’s condition and the state of the environment. Physiological data were collected using an Empatica E4 wristband equipped with sensors measuring blood volume pulse (BVP) and the galvanic skin response (GSR), which allow deriving the heart rate and arousal (Bach, Friston, & Dolan, 2010; Benedek & Kaernbach, 2010). The E4 is also equipped with an infrared thermopile sensory which reads peripheral skin temperature and a 3-axis accelerometer capturing arm motion. The simulator was equipped with a camera recording student’s face for later facial expression analysis to measure valence. The E4 and facial expressions had been used previously in research settings to assess some of the pilots’ affective states (Lawrynczyk, Chaouachi, & Lajoie, 2017).

The data generated by the C172 data was used to verify the solution approach with regards to its ability to handle substantially different types of training data (as compared to data logged by strategy games) and to verify the implementation of modular analyzers (which are a vital component of the overall solution approach).

4.3 Operational Validation

A human subject study with civilian aviators was planned and executed to validate the expected operational benefits of introducing automated human performance analysis and scenario recommendations into the training cycle. More than 20 pilots were recruited from three local flight schools, Hampton University and ODU, including several military pilots. These pilots were brought to VMASC, introduced to the research, and explained the testing plans and goals. Pilots were given two hours of instruction and unanalyzed flight time to get oriented to the testing simulator. No data was collected during this time.

A data collection computer was set up and connected to the C172 simulator to download the simulation data from each run. Pilots were briefed on the testing T&R matrix, how the matrix will track their "readiness" in the simulator, and initial runs were conducted on each pilot. These initial runs tested the administrative, operational, and technical actions required to initiate a new T&R readiness program for each pilot, perform that training in the C172 simulator, and update their T&R matrix.

The authors researched aviation-relevant MOPs that could be tested and measured in the C172 simulator, then developed a list of skills that could be used to populate a testing T&R matrix. Once those measurable skills were identified, our team validated the simulator's ability to present realistic testing scenarios where those skills could be evaluated. Once test runs were validated, the team ensured data outputs allowed to identify those tested parameters. Based on the performance of the aircraft simulation when being flown by experienced C172 instructor pilots, optimal and minimum acceptable measures of performance were established for test subjects (see Table 1).

Table 1. T&R Skills and Associated Measures of Performance.

Skills	Standard	Perfect	Objective	Threshold
1	Maintain course (climbout) - day/VFR	0 deg. deviation	+/- 3 degrees	+/- 5 degrees
2	Maintain course (climbout) - IFR	0 deg. deviation	+/- 5 degrees	+/- 10 degrees
3	Maintain course (climbout) - night	0 deg. deviation	+/- 5 degrees	+/- 10 degrees
4	Maintain airspeed (climbout) - day/VFR - 74 KIAS	74 KIAS	+/- 5 knots	+/- 10 knots
5	Maintain airspeed (climbout) - IFR - 74 KIAS	74 KIAS	+/- 10 knots	+/- 15 knots
6	Maintain airspeed (climbout) - night - 74 KIAS	74 KIAS	+/- 10 knots	+/- 15 knots
7	Maintain airspeed (level) - day/VFR	0 knots deviation	+/- 5 knots	+/- 10 knots
8	Maintain airspeed (level) - IFR	0 knots deviation	+/- 10 knots	+/- 15 knots
9	Maintain airspeed (level) - night	0 knots deviation	+/- 10 knots	+/- 15 knots
10	Maintain altitude - day/VFR	0 feet deviation	+/- 50 feet	+/- 100 feet
11	Maintain altitude - IFR	0 feet deviation	+/- 150 feet	+/- 250 feet
12	Maintain altitude - night	0 feet deviation	+/- 100 feet	+/- 200 feet
13	Turn at fixed rate - day/VFR	3 degrees per second	+/- 2 degrees per second	+/- 4 degrees per second
14	Turn at fixed rate - IFR	3 degrees per second	+/- 6 degrees per second	+/- 9 degrees per second
15	Turn at fixed rate - night	3 degrees per second	+/- 4 degrees per second	+/- 6 degrees per second
16	Hold VSI (climbing on departure) - day/VFR	1000 feet per minute up	+/- 100 feet per minute up	+/- 200 feet per minute up
17	Hold VSI (climbing on departure) - IFR	1000 feet per minute up	+/- 250 feet per minute up	+/- 400 feet per minute up
18	Hold VSI (climbing on departure) - night	1000 feet per minute up	+/- 200 feet per minute up	+/- 300 feet per minute up
19	Hold VSI (climbing - maneuvering) - day/VFR	500 feet per minute up	+/- 100 feet per minute up	+/- 200 feet per minute up
20	Hold VSI (climbing - maneuvering) - IFR	500 feet per minute up	+/- 250 feet per minute up	+/- 400 feet per minute up
21	Hold VSI (climbing - maneuvering) - night	500 feet per minute up	+/- 200 feet per minute up	+/- 300 feet per minute up
22	Hold VSI (descending - maneuvering) - day/VFR	500 feet per minute down	+/- 100 feet per minute down	+/- 200 feet per minute down
23	Hold VSI (descending - maneuvering) - IFR	500 feet per minute down	+/- 250 feet per minute down	+/- 400 feet per minute down
24	Hold VSI (descending - maneuvering) - night	500 feet per minute down	+/- 200 feet per minute down	+/- 300 feet per minute down
25	Hold VSI (descending) - day/VFR	500 feet per minute down	+/- 100 feet per minute down	+/- 200 feet per minute down
26	Hold VSI (descending) - IFR	500 feet per minute down	+/- 250 feet per minute down	+/- 400 feet per minute down
27	Hold VSI (descending) - night	500 feet per minute down	+/- 200 feet per minute down	+/- 300 feet per minute down
28	Maintain course (level flight) - day/VFR	0 deg. deviation	+/- 3 degrees	+/- 5 degrees
29	Maintain course (level flight) - IFR	0 deg. deviation	+/- 5 degrees	+/- 10 degrees
30	Maintain course (level flight) - night	0 deg. deviation	+/- 5 degrees	+/- 10 degrees

Skills required to fly a C172 and that can be evaluated in the C172 simulator were used to develop a baseline aviator training requirement in the form of a T&R matrix (see Figure 7). Numerous mini sequences were developed to test each of the skills in this new T&R matrix. Those sequences were pieced together to create training scenarios. Six initial master training scenarios, with variations for weather and time of day, were developed with periodicities that ensured all skills were trained regularly. Figure 8 shows an example scenario description provided to the pilots (plus accompanying flight charts etc.).

Mission Essential Tasks	MOB101 Daytime flight	MOB102 Nighttime flight	MOB103 Instrument flight	MOB104 C-172 Emergency Procedures	MOB201 Visual practice pattern (at altitude)	MOB202 Instrument approach	MOB203 Landings	MOB204 Basic Level Turns - Daytime	MOB205 Low-altitude flight	MOB206 Basic Level Turns - Nighttime	MOB207 Basic Climbs and Descents	MOB208 Basic Three-Dimensional Maneuvering	MOB209 Basic Climbs and Descents - Nighttime	MOB210 Basic Level Turns - Bad weather	MOB211 Basic Pattern Flying	MOB212 Basic Pattern Flying - Nighttime
1) Fly straight and level	X	X	X	X					X					X	X	X
2) Turn at a fixed rate	X	X	X		X	X	X	X	X	X		X		X	X	X
3) Climb at a fixed rate	X										X	X	X		X	X
4) Descend at a fixed rate											X	X	X		X	X
5) Climb and turn at a fixed rate						X					X	X	X			
6) Descend and turn at a fixed rate							X					X	X		X	X
7) Level off on altitude after climb	X	X			X		X				X	X	X		X	X
8) Level off on altitude after descent					X		X				X	X	X		X	X
9) Fly visual landing pattern					X		X								X	X
10) Fly glidepath					X	X	X								X	X
11) Fly lineup					X	X	X								X	X
12) Land airplane	X			X		X	X	X							X	X
13) Fly airplane at night		X								X			X			X
14) Land airplane at night		X					X			X			X			X
15) No flap takeoff	X	X														
16) Full flap takeoff	X	X														
17) Ground taxi	X	X					X									
Periodicity	30	60	60	365	60	60	30	60	60	60	60	120	120	60	60	120

Scenario 1 Basic VFR flight	Scenario 2 Airways Navigation	Scenario 3 Non-precision approaches	Scenario 4 Emergency Procedures	Scenario 5 Airways Navigation (night)	Scenario 6 Non-precision approaches (night)	Scenario 7 Basic Aerobatic Flight
101	101	101	101	102	102	101
104	103	103	103	202	202	204
201	104	104	104	205	203	205
202	204	202	203	209	206	207
203	205	203	211		209	208
204	206	207			212	
206	207	210				
207	210					
210	211					

Figure 7. Initial Cessna 172 T&R Matrix.

<u>Scenario 2:</u> IFR airway navigation flight from KNGU to KORF, ILS approach at KORF.

<u>Description</u>: Perform no-flap takeoff from runway 10 at NGU; Turn left to a heading of 090 and climb to 1,000'; Turn left to intercept LUFSY and then climb to 1,600' MSL; Fly through JHALL on a heading of 077. Accelerate to and maintain 110 KIAS; Fly outbound on a heading of 077 and make a left-hand standard rate turn while maintaining 1,600' to intercept the ORF 047 radial and fly the final approach course of the ILS approach; Cross LUFSY at 1,600' inbound and descend per the approach.

Figure 8. Example scenario description.

The authors continued to refine the internally generated C172 T&R matrix to align with a standard Navy T&R Matrix. Early testing showed our initial matrix had skills in the area generally reserved for Mission Essential Tasks (left side of T&R matrix). This required a rewrite of the matrix to move the tested skills out of the matrix and make them internal parts of the Flight Tasks. Mission Essential Tasks (METs), skills required to meet our 'mission' (be a good Cessna pilot), were then put on the left side of the matrix. These tasks, such as IFR navigation flight, can be broken down into numerous skills required to perform those tasks successfully (i.e., maintain speed control, maintain the angle of bank control, etc.). The Flight Tasks (across the top of the matrix) were then paired with METs to show where those METs could be accomplished. Flight Tasks were then broken down into what skills are required for their successful accomplishment. The Flight Tasks were pieced together to develop training scenarios. The master training scenarios, with variations for weather and time of day, were then given periodicities to ensure all skills were trained regularly. Figure 9 shows the refined T&R matrix.

	Flight Tasks																								
	Fly straight and level - day/VFR	Fly straight and level - night	Fly straight and level - IFR	Turn at a fixed rate - day/VFR	Turn at a fixed rate - night	Turn at a fixed rate - IFR	Climb at a fixed rate - day/VFR	Climb at a fixed rate - night	Climb at a fixed rate - IFR	Descend at a fixed rate - day/VFR	Descend at a fixed rate - night	Descend at a fixed rate - IFR	Climb and turn at a fixed rate	Descend and turn at a fixed rate	Level off on altitude after climb	Level off on altitude after descent	Fly glidepath	Fly lineup	Fly non-precision approach - day	Fly non-precision approach - night	Fly non-precision approach - IFR	Fly precision approach - day	Fly precision approach - night	Fly precision approach - IFR	Fly airplane at night
Daytime Navigation Flight	X			X						X			X	X	X	X	X	X	X			X			
Nightime Navigation Flight		X			X			X	X		X		X	X	X	X	X	X		X			X		X
IFR Navigation Flight			X			X	X		X			X	X	X	X	X	X	X			X			X	
Aerobatic maneuvering flight				X	X	X	X	X							X										X
Emergency Procedures	X			X																					X
Airwyays Navigation Situational	X	X	X	X	X	X	X	X	X	X	X	X							X	X	X				X
Mountain Flying	X			X			X			X			X	X	X	X			X						
Periodicity	10	20	25	10	20	25	10	20	25	10	20	25	10	15	20	15	15	15	10	20	25	10	20	25	20

Figure 9. Refined Cessna 172 T&R Matrix.

Linkages between T&R matrix Flight Tasks, skills, MOPs and scenarios were established to help the system learn:

- What skills are tested under each Flight Task
- What MOPs are established for each skill
- What Flight Tasks are combined into the current scenarios

This design will allow the system to learn how Flight Tasks and skills are related to each other, and to adjust future scenario recommendations to emphasize skills that need more focus and training.

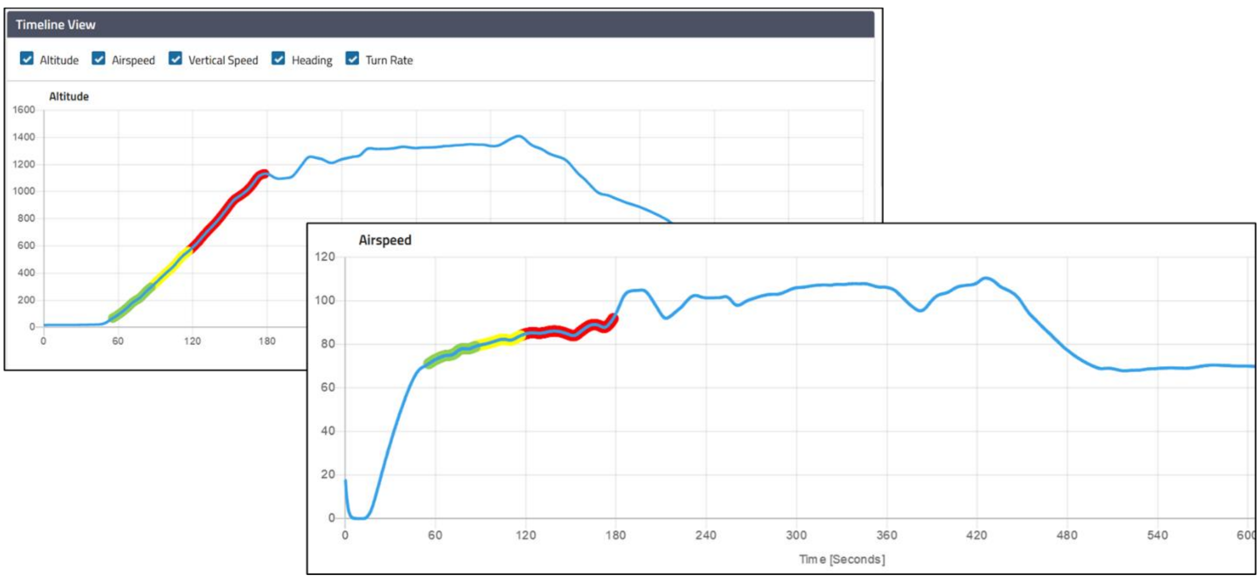

Figure 10. Example Visualization of Aircrew Performance.

Once the prototype solution, the C172 simulator, the instructional material and T&R matrix were prepared, a 6-month training plan was executed. Finally, 20 aircrews were trained in 100+ simulator events with a total of 700+ scenario executions. During the process, 7+ GB of training data were collected. Figure 10 shows examples visualizations of aircrew performance, including the actual data points and the color-coded analysis results (green= objective met, yellow= threshold met, red= unsatisfactory performance). Multiple visualizations are available to give SMEs and aircrew insight into each flight. The example in Figure 10 shows the altitude and airspeed during takeoff.

4.4 Analysis And Lessons Learned

4.4.1 Technical Aspects

The prototype system used for concept validation has been successfully tested and demonstrated the feasibility of the concept design and overall system architecture. Specifically, the concept validation has shown the system's ability to handle four distinct training environment data sets: Gaming data, C172 flight training device data, Navy Helicopter Sea Combat (HSC) flight simulator data, and the U.S. Army's Aviator Training Next (ATN) system data.

Although each training system provides training data in a slightly different format, many commonalities have been identified. They are likely the result of focusing on a specific class of simulators, flight simulators in our case. Individual data loading processes (including, data cleaning, data preparation etc.) must cope with individual deviations in data types, data aggregation and data resolution.

4.4.2 Data Analysis

The selected approach with specific analyzers for individual MOPs was demonstrated successfully. For organizational reasons, the concept validation phase focused primarily on automated assessment of basic flying skills. It was found that these skills are comparatively easy to assess and it is expected that higher-level skills (like properly engaging an enemy aircraft) are more challenging to determine. The underlying concept of using modular analyzers should still be applicable, albeit it is expected that the individual analyzers become more complex and may require more sophisticated data analysis techniques.

One specific identified issue is that the analyzers had to know when to start evaluating a particular skill. Initially, the system was told what training was occurring and when to start and stop ‘looking’ at specific parameters. It worked very well for pre-established, recurring training events. On the other hand, fleet training tends to have less structure. It will require the system to be able to identify when specific skill training has started and stopped. To illustrate, if an aircraft’s location is at a particular field, altitude is ground level, and Vertical Speed changes from 0 to >100 ft/min for 3 seconds, the climb skill is triggered. The climb rate will be monitored against established norms. The end trigger will occur when Vertical Speed changes to <100 ft/min for 3 seconds.

4.4.3 Operational Benefits

The prototype system's capability to make individual training recommendations (based on currency and proficiency) could successfully be demonstrated. Unfortunately, extensive validation against a control group (that uses the traditional approach to training) could not be executed due to resource constraints. The students participating in the study, various active Training and Readiness Officers and aircrew emphasized that automated trainee assessment and individualized training recommendations are of extraordinary value. The preliminary results indicate that such a system can successfully address both key challenges (see Section 1.3), specifically automating training data analysis, reducing SME workload, and enabling trainees to self-assess their performance.

5. CONCLUSIONS AND NEXT STEPS

To overcome two critical challenges of the current training (specifically, dependence on manual tasks and limited availability of SMEs), the authors have developed a data analytics approach that automates trainee performance assessment, thereby generating objective MOPs for each trainee and provides recommendations for best-suited future training.

The overall concept and system design were successfully evaluated using a prototype system implementation tested with different training data sets (simulators, volumes, and formats). A fully functioning user interface to support test and evaluation was developed for the initial testing runs. The prototype system was configured to accept testing data regularly and accounts for all pilots were created. Initial data from each pilot, collected during initial training runs, was entered to provide the system with a skill baseline for each pilot.

The prototype system successfully demonstrated the viability of the solution approach, i.e., to analyze basic training data, automatically identify skills being trained and analyze MOPs. It then presents recommendations for future training based on previous training data and periodicity requirements for each Flight Task.

Once sufficient flight simulator data is collected from the pilots participating in the human subject study, selected analyzers will include models resulting from machine learning to replace the heuristical approach where it would be more effective. Proper labels, such as “take off,” “landing,” and “stress,” will be applied manually and (semi-) automatically. The AI-based analysis will also allow identifying emerging MOPs from a large amount of data currently being collected.

Further research and development will be required to allow for observer-based validation of the skills being mastered and real-time analysis of trainee performance and scenario adjustment. There may also be a need for the observer to validate the outcome evaluation. The current prototype system may or may not identify a skill being trained appropriately. An observer can 'guide' the system to learn the task or skills being taught. The vital aspect of this guidance is that non-SMEs or the trainee post-event can do it.

ACKNOWLEDGEMENTS

This work was sponsored by the Naval Air Warfare Center Training Systems Division (NAWCTSD) under a Small Business Technology Transfer (STTR) effort.

REFERENCES

Bach, D. R., Friston, K. J., & Dolan, R. J. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *International journal of psychophysiology*, *76*(1), 52-55.

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, *190*(1), 80-91.

Cohen, R. S., Chandler, N., Efron, S., Frederick, B., Han, E., Klein, K., . . . Shokh, Y. (2020). *The Future of Warfare in 2030: Project Overview and Conclusions*: RAND Corporation.

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review, 1*(2), 293-314. doi:10.1093/nsr/nwt032

Freeman, J., & Zachary, W. (2018). Intelligent tutoring for team training: Lessons learned from US military research. In *Building Intelligent Tutoring Systems for Teams*: Emerald Publishing Limited.

Hodson, D. D., & Hill, R. R. (2014). The art and science of live, virtual, and constructive simulation for test and analysis. *The Journal of Defense Modeling and Simulation, 11*(2), 77-89.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society, 1*(1), 2053951714528481.

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proc. VLDB Endow., 5*(12), 2032–2033. doi:10.14778/2367502.2367572

Lawrynczyk, A., Chaouachi, M., & Lajoie, S. P. (2017). *Multimodal Assessment of Pilots' Affective States using Psychophysiological Sensor Signals and Facial Recognition Analysis*. Paper presented at the The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Miller, De Wolfe H. (2018). Warfighting, people and readiness. *Naval Aviation News*, *March 21, 2018*. Retrieved October 10, 2018, from http://navalaviationnews.navylive.dodlive.mil/2018/03/21/warfighting-people-and-readiness

Sayler, K. M. (2020). *Emerging Military Technologies: Background and Issues for Congress*. Congressional Research Service Washington United States Retrieved from https://fas.org/sgp/crs/natsec/R46458.pdf

Song, X., Wu, Y., Ma, Y., Cui, Y., & Gong, G. (2015). Military Simulation Big Data: Background, State of the Art, and Challenges. *Mathematical Problems in Engineering, 2015*, 298356. doi:10.1155/2015/298356

Synnaeve, Gabriel, Bessiere, Pierre. A Dataset for StarCraft A.I. & an Example of Armies Clustering. AAAI Workshop - Technical Report. 2012. https://arxiv.org/abs/1211.4552