Performance Assessment Using Individual Skills Linked to Mission Outcomes

Brad Gilroy
2 Circle Consulting, Inc.
Linthicum, Maryland
bgilroy@2circleinc.com

Dave Harris
2 Circle Consulting, Inc.
Smithfield, North Carolina
dharris@2circleinc.com

ABSTRACT

Traditional performance assessment techniques utilize a top-down approach focused on mission outcomes (i.e., "what" occurred). These techniques often fail to uncover "why" the outcomes occurred, which is critical information for training and resourcing decision makers. An alternative approach that focuses on the factors that directly affect mission outcomes (i.e., individual skill execution coupled with systems reliability) has traditionally been deemed "impossible" due to the fidelity of data required. However, in 2015, a team of former TOPGUN instructors overcame challenges (e.g., disparate data sources, historically subjective/self-reported assessments, etc.) to develop a methodology that links the performance demonstrated by aircrew at the tactical level of execution to team/mission outcomes during training events. Over the last five years, this new approach to data collection and analysis has been developed into an automated toolset and applied to hundreds of aircrew training events. In addition to providing invaluable feedback to the aircrew, the quantifiable data collected during these events provide objective insight into the individual skill and aircraft systems deficiencies that have the greatest impact on mission outcomes. Through the application of machine learning techniques, the data is also used to forecast future outcomes and derive standards for individual performance to meet combat objectives. Validation of this approach comes from the fact that the Naval Aviation community now relies upon the resulting data and analysis as the means to solve training and proficiency-related questions to include syllabi updates and training resource investments. Although initially demonstrated in the context of F/A-18 air warfare, the process has been extended across multiple platforms (e.g., E-2C/D, EA-18G, F-35, etc.), missions, and training environments (i.e., live, virtual/constructive, and distributed). The purpose of this paper is to highlight the skills-based approach to performance assessment, once deemed impossible to achieve, that is now being implemented by the United States Navy.

ABOUT THE AUTHORS

Brad Gilroy is a former F/A-18 pilot and graduate of the United States Navy Fighter Weapons School (TOPGUN). In 2015, Mr. Gilroy helped develop the Aircrew Performance Measurement (APM) data collection process for what would become the Training Analysis Project (TAP). He is one of only a few APM specialists qualified to collect, analyze, and report on F/A-18, EA-18G, E-2, and F-35 aircrew performance and has been involved in numerous aircrew proficiency studies across the Naval Aviation Enterprise (NAE). Mr. Gilroy is a senior analyst and group manager for 2 Circle Consulting, Inc. as well as the project lead for the development, marketing, and sales of the Reconstruction and Assessment of Proficiency in an Integrated Debrief (RAPID) software, which he co-invented in 2018. RAPID effectively automates the APM process for use across multiple Department of Defense (DoD) platforms as well as commercial applications. Mr. Gilroy is a 1999 graduate of the United States Naval Academy.

Dave Harris is a senior analyst and division manager for 2 Circle Consulting, Inc. He is a former F-14 and F/A-18 pilot and two-time instructor on the TOPGUN staff. Mr. Harris was the lead training and technical expert during the development of the APM data collection process currently used by TAP. He is responsible for overseeing the Fleet Support Division of 2 Circle, which includes the APM data collection team as well as all studies and Fleet support functions serviced by APM data. In 2018, Mr. Harris co-invented the RAPID software with Mr. Gilroy. Mr. Harris is a 1996 graduate of North Carolina State University.

Performance Assessment Using Individual Skills Linked to Mission Outcomes

Brad Gilroy
2 Circle Consulting, Inc.
Linthicum, Maryland
bgilroy@2circleinc.com

Dave Harris
2 Circle Consulting, Inc.
Smithfield, North Carolina
dharris@2circleinc.com

INTRODUCTION

Traditional performance assessment techniques utilize a top-down approach focused on mission outcomes (i.e., "what" occurred) to determine the probabilistic risk associated with the overall system. This event-based approach rarely answers questions about the underlying factors that shape mission outcomes. Unfortunately, when developing syllabi to address individual performance deficiencies, just knowing "what" occurred is not enough. When making training or resourcing decisions, syllabi developers also need a thorough understanding of "why" those outcomes occurred. Using the F/A-18 community as an example, when conducting simulated air-to-air engagements, it is important to know the likelihood that a fighter will "die" in training at the hands of a threat. However, to develop syllabi that accurately and efficiently address deficiencies in aircrew performance, it is also important to understand the conditions that existed when the fighter died in order to truly understand the cause of that "death."

Comprehensive analysis of the underlying factors that shape mission outcomes requires a higher level of data complexity. As a result, it falls into a category of analysis that is traditionally labeled as "resource intensive" and is often assumed to be too cost prohibitive to apply. The Capabilities-Based Training and Readiness Matrix (CBTRM) currently utilized by the military is an event-based system that assigns and tracks prescribed training events, but has no means to measure and report the proficiency of the aircrew in the skills and tasks meant to be developed during the training events. Instead, CBTRM tracks warfighting readiness as a function of resource expenditures. Under this system, neither individual proficiency nor unit performance is factored into assessments of unit readiness. As a result, the system lacks the quantifiable data to answer questions regarding the proficiency of any aircrew at the assigned skills, tasks, and missions.

In 2015, a team of former United States Navy Fighter Weapons School (TOPGUN) graduates and instructors, in support of the Naval Air System Command (NAVAIR) and the Commander Naval Air Forces (CNAF), developed a methodology that, unlike CBTRM, quantified the performance of individual F/A-18 aircrew and aircraft in a way that directly linked to overall mission outcomes. This new, bottom-up approach to data collection and performance assessment focuses on individual-level execution and system-level capability to determine whether, and why, the mission was a success or failure. Over the past five years, this performance assessment methodology has been applied and validated during hundreds of F/A-18 training events. New automated tools have been developed to provide more efficient assessment and to enable application of the process across multiple platforms and mission sets. This new system, which is scalable and adaptable to a wide range of performance assessment programs, is improving training and resourcing decisions across the Naval Aviation Enterprise (NAE).

Capabilities-Based Training and Readiness Matrix (CBTRM)

For almost 20 years, the Department of Defense (DoD) has utilized an events-based approach to track readiness. Every month, units provide a summary of the events that were executed and resources that were utilized. Unfortunately, this approach does not provide quantitative feedback to NAE leadership on the resulting performance demonstrated by the warfighter during the execution of each task. Without quantifiable data, it is impossible for the NAE, or any organization for that matter, to understand:

- The amount of training required to achieve proficiency objectives
- The resources that need to be applied to an individual or unit to achieve the specified objectives
- The aircrew that need additional/extra training to achieve proficiency objectives
- The skills that require focus in follow-on training events
- The degree of assumed/accepted risk given the current levels of proficiency

Because the current CBTRM system does not accurately measure performance, track individual proficiency, nor assess warfighting capability (i.e., provide quantitative feedback on individual/unit performance), the DoD has no other option but to apply equal resources to all individuals/units in a boilerplate fashion. For example, the Commanding Officer of an F/A-18 squadron, with considerably more experience in the aircraft, is provided the same resources/training as the most junior aviator in the squadron, who has considerably less experience/proficiency. This also occurs at the unit level where squadrons are given the same training, regardless of their proficiency and resourcing. This boilerplate approach to training/resourcing is inefficient and, in today's resource-constrained environment, an improved system is needed to optimize the application of resources. This can only be accomplished through a data-driven, tailored approach to training.

This is not to say that training in the DoD is conducted without any performance assessment whatsoever. Certainly, some level of performance assessment is being conducted by the individual warfighters and instructors during post-event debriefs to develop lessons learned for future employment. However, the fact remains that the current system of readiness reporting does not institutionalize performance assessment nor provide the tools to standardize the way assessments are conducted. As a result, the warfighters are often left to conduct their own subjective assessments of their performance through a complex post-event debrief process that involves multiple tools/sources of data and is absent of any comparisons to known standards that link to mission success in combat.

Training Analysis Project (TAP)

Recognizing the deficiencies associated with CBTRM, particularly with regards to the air warfare mission sets, the Office of the Chief of Naval Operations (OPNAV) and CNAF asked two critical questions:

- What is our combat capability to meet the threat today?
- What level of training is required to meet the threat of tomorrow?

In response to this challenge, NAVAIR established the Training Analysis Project (TAP) in 2015. The TAP team combines the expertise of operational warfighting/training subject matter experts from 2 Circle Consulting, analytical and mission modeling support from Johns Hopkins University – Applied Physics Lab (JHU/APL), operational research psychologists from the Naval Air Warfare Center Training Systems Division (NAWCTSD), and systems analysts from NAVAIR's Aviation Readiness and Resource Analysis Department (AIR-6.8). The mission of TAP is to create a data-driven, performance-based methodology for evaluating individual and unit-level readiness to meet the needs of today while also providing prescriptive recommendations to ensure enduring capability in the future. Over the past five years, TAP has been developed from a small group of analysts focused on a single platform and mission set (F/A-18 air warfare) into the NAE's only source for objective data on aircrew performance measured against the warfighting standards derived from the Navy Mission Essential Tasks (NMETs), as well as the training requirements derived from the Operational Plans (OPLANs) of the Naval Component Commanders (NCCs). The data collected by TAP enables analysis on the performance capabilities of the Fleet, calculation of the risk associated with identified deficiencies, and the actionable information required to optimize training and resourcing.

Military personnel are often placed in difficult situations that cannot be solved with a single solution. As a result, the syllabi that are used to train military personnel often put them in a variety of settings that involve active learning and force the individual warfighters to apply the knowledge to solve problems and perform skills/tasks. The challenge for TAP was to develop a performance assessment methodology that could meet these demands and be applied across all platforms, missions, and training environments (e.g., live, virtual, and constructive). This flexibility, however, resulted in a framework for conducting the process that is extensible beyond just DoD platforms and missions (i.e., not just F/A-18 air warfare but commercial applications as well).

METHODOLOGY

Most top-down performance assessment techniques rely on the collection of all available data followed by analysis post-collection to determine relevance. Unfortunately, only some of the data collected under this approach are truly relevant. TAP's new approach required a much more efficient methodology for collecting high-fidelity data on individual skill execution. As a result, the TAP team attacked the problem in reverse. Rather than collecting, storing, and reporting all available data, TAP conducted a targeted analysis of the data sources used by the F/A-18 community to determine relevance first, before developing the collection processes, automated performance assessment tools, data storage requirements, and reporting methods. This informed approach streamlined the data collection and data storage requirements and ensured that the TAP data/analysis only focused on metrics that were proven to be traceable to desired outcomes.

Since its inception, the TAP data collection process has been applied to hundreds of F/A-18 training events. The results provided unprecedented insight into the individual skills and aircraft system deficiencies that have the greatest impact on F/A-18 mission outcomes. The initial success validated the need to analyze the metrics first and put them into the correct context before instituting advanced processes for collection, analysis, and storage. Armed with this analysis, the next logical step for TAP was to develop software tools to automate the performance assessment process. The Reconstruction and Assessment of Proficiency in an Integrated Debrief (RAPID) software began development in 2019 for Fleet implementation with the TAP team starting in 2021. The software incorporates ground-breaking applications of Artificial Intelligence/Machine Learning (AI/ML) techniques that enhance the quality, quantity, and utility of the performance data and supports statistical analysis of the dataset in a single application. RAPID incorporates all the processes, methodologies, and lessons learned from TAP and provides the means for Fleet aircrew to perform these functions on every training event, enabling a performance-based approach to training that can replace the inefficient event-based approach that exists today.

Aircrew Performance Measurement (APM) Process

The current TAP methodology is extensible to many applications; however, when applied to air platforms such as F/A-18, it is also known as the Aircrew Performance Measurement (APM) process. The APM process starts six months before an air wing begins their Optimized Fleet Response Plan (OFRP) training. During this time, TAP conducts a Resourcing and Readiness Assessment (RRA) of the air wing's resources up to that point. The RRA includes data and analysis on the following:

- Air wing composition
- Flight hour execution vs. allotment
- Aircrew manning and qualification levels vs. allotment
- Maintainer manning and qualification levels vs. allotment
- Systems availability/health
- Support equipment availability/health
- Expected syllabi (i.e., schedule of events, training objectives, etc.)
- Actual performance at the individual, squadron, and air wing levels in each syllabus
- Projected performance in subsequent syllabi based on current vs potential resourcing and readiness

Individually, each component of the RRA is unremarkable. However, when compiled together into a consolidated report, it provides NAE leadership with a complete picture of the air wing's current and potential capabilities. Compilation of the data 6 months prior to the start of training affords NAE leadership with ample time to address deficiencies before the start of training. New RRAs are compiled 3-months prior and at the start of the OFRP to provide updates on the air wing's resourcing status.

Additionally, APM data collected during previous exercises is used to provide the air wing with valuable insight into the upcoming training. At the start of OFRP training, each air wing is given a series of briefings on Fleet-wide trends, common errors, recommended areas of focus before and during training, etc. These briefings are supported by a data-driven virtual training syllabus to develop a required minimum level of proficiency in skills that the APM data has shown to be critical for mission success before any live flight training. Once OFRP training begins, the APM data collection team also provides daily and weekly reports on individual, unit, and air wing performance as well as trend information to air wing and squadron leadership while executing the OFRP syllabi to inform real-time training decisions and remediation.

The APM data collection focuses on five categories of data, shown in Table 1. The categories are designed to capture all the potential aspects of demonstrated performance by identifying the aircrew, aircraft configuration, training/mission objectives, executed Tactics, Techniques, and Procedures (TTPs), applied tasks/skills, threats, range/ training environment, etc.

Table 1. APM Data Categories and Associated Metrics/Parameters

Category	Example Metrics/Parameters
Mission Outcomes	Kills vs. Deaths (i.e., kill ratio)
	Mission-dependent success criteria
Aircrew Performance	Individual aircrew tactical errors assessed against the standards outlined by NAE policy through measures of performance/measures of effectiveness and TOPGUN standards for tactical execution
Systems Availability/Health	Aircraft configuration
	Systems status
Aircrew Biographical Information	Crew position
	Military rank
	Flight hours
	Previous assignments (i.e., career experience)
	Qualifications
	History of past training exercises
Event Details	Mission/Training objectives
	Selected tactics
	Threat details
	Simulated loadout
	Range/Training Environment Details

In the air warfare mission set, aircraft capability is a significant contributor to warfighting effectiveness. A technological advantage over an adversary improves the chances for mission success. However, maximum aircraft capability can only be achieved when the aircraft is configured with all of its systems and those systems are maintained in optimal condition. While the capability of each system on the F/A-18 is well documented, the availability and health of each system vary from one event to the next. Individual system availability (i.e., was the system installed?) and health (i.e., was the system functioning properly?) are documented as part of the APM process to quantify the impact that each system has on mission outcomes. The following model is used to demonstrate this concept for illustrative purposes only:

Systems Reliability = Systems Capability x Systems Availability x Systems Health

However, the aircraft is only one component of mission success. Aircrew are also critical to the success of any mission and must be considered when analyzing warfighting effectiveness. In the F/A-18 community, TOPGUN is responsible for developing and prescribing the TTPs that Fleet aircrew use to be lethal and survivable in combat scenarios. To quantify the impact that aircrew have on mission outcomes, the APM process evaluates each individual based on their ability to perform those TTPs as prescribed. The following model is used to demonstrate this concept for illustrative purposes only:

Aircrew Reliability = Aircrew Performance x TTPs

Using the concepts described above, the APM data on aircrew performance and systems availability/health are used to quantify the effectiveness of the entire system (i.e., man and machine) and provide context to the mission outcomes. The fidelity of the data captured by the APM process is sufficient to trace every documented mission outcome (e.g., simulated fighter death in training) back to either a human error (e.g., aircrew incorrectly applying defensive TTPs), a systems error (e.g., missing or degraded defensive system), or a combination of the two. In doing so, the APM process fulfills the NAEs desire to have a data-driven methodology to assess risk. This is accomplished using the following model, which demonstrates how warfighting effectiveness can be calculated for use in various warfare analysis and modeling efforts by applying quantifiable data on aircrew and aircraft reliability:

Warfighting Effectiveness = Aircrew Reliability x Systems Reliability

The model illustrates that warfighting effectiveness increases (i.e., risk is minimized) when the right systems are put in the hands of the right aircrew so they can be employed to their maximum potential. In addition to providing the

NAE with a methodology for assessing risk, the model also provides the NAE with a way to “reverse engineer” the training and systems requirements (i.e., investment decisions) that will ultimately achieve the desired outcomes in combat.

When collecting performance data, the APM process also calls for the collection of data that provides context about the aircrew who were involved in the training. Recognizing that not all aircrew perform equally, aircrew biographical data is collected and paired with individual performance data to support analysis that identifies which individual traits or training backgrounds result in higher performance. This analysis quantifies the Return on Investment (ROI) associated with individual training tracks. For example, in the F/A-18 community, analysis of aircrew biographical data confirmed that the “quality” of training is far more important than the “quantity” of training. Individual aircrew with significantly more experience (i.e., flight hours) were outperformed in almost every metric by individuals with less experience whose background included recent, focused, high-fidelity training events.

The APM process also calls for collection of data that provides context about the training event itself. Recognizing that individual performance can fluctuate from one event to the next, details are collected on the training event (e.g., mission/training objectives, airspace/range details, adversary force presentation/capability, etc.) to identify the outside factors that have the biggest effect on individual performance and mission outcomes. This quantifies the ROI associated with various training techniques. It also provides the means to explain the relative difficulty of one event compared to another as a way to explain why mission outcomes may fluctuate even when individual skill performance remains the same.

Of note, all performance data collected by TAP is marked with the source(s) of the data to qualify the data quality. TAP only publishes data that is captured using one or more high-quality data sources. All data that is provided by the individual aircrew is retained; however, due to the low reliability of self-reported data, it is not reported by TAP or used to inform training/resourcing decisions by the NAE.

Automating the APM Process

The process used to collect data when TAP in 2015 was a lengthy process that required primarily manual derivation and compilation of data points from the multiple, disparate sources of F/A-18 performance data. At that time, the project depended heavily on the manpower, credibility, diligence, and proficiency of the APM data collectors. To ensure credibility, multiple data collectors were required to be on site to observe a single element of fighters, and all data collectors were required to be TOPGUN graduates (i.e., certified Strike Fighter Tactics Instructors). To ensure consistency, data collectors were also required to retain an expert level of knowledge in air warfare training programs, as well as aircraft and weapon hardware, software, and TTPs via standardization training. Finally, TAP data collectors were required to maintain an expert level of proficiency in operating all the training systems that provided the disparate sources of aircrew performance data.

To overcome these challenges and increase the effectiveness of the APM process, new tools needed to be developed that would pre-process, fuse, and store data from multiple sources and logically walk any user, regardless of their proficiency with the data or data sources, through the APM data collection process. These new tools provide the ability to automate the collection, assessment, and reporting of aircrew/aircraft performance metrics. Additionally, since APM data collectors use the same data sources as Fleet aircrew, these new tools essentially put the APM process directly in the hands of Fleet aircrew and provide them with the means to execute a reliable and accurate assessment of their performance during the post-event debrief. The goal of automating the APM process was to generate objective, quantifiable data while also improving combat capability by optimizing the learning process applied to each training event.

The automated APM tools provide aircrew and data collectors with a single, common interface that facilitates the APM process for all platforms, missions, and training environments (e.g., live, virtual, and constructive). By consolidating all the sources of data into a single one-stop-shop, the software eliminates the segmented approach currently used to reconstruct, analyze, and assess performance and replaces it with a comprehensive debrief methodology where all information is co-located and can be assessed without interruption. Rather than relying on the user to extract pertinent information themselves, the tool actively walks them through the APM process, providing cues that are extracted directly from the digital sources of F/A-18 data to ensure that valuable learning points are not missed. Bottom line, automation reduces the workload on data collectors, improves the consistency of the data, and amplifies the speed and reliability of feedback provided to aircrew and NAE decision makers, enabling Fleet-wide implementation of a performance-based approach to training and resource management with assessments made against objective, traceable standards.

Using Artificial Intelligence/Machine Learning (AI/ML) to Derive Data-Driven Performance Standards

Data without context is useless. The only way to provide context to individual performance is to compare the data to defined standards. For example, the APM data quantifies the F/A-18 community's ability to correctly employ AIM-120 missiles, which is a critical skill. That number (i.e., the percentage of correct/incorrect AIM-120 employments), by itself, is interesting but does not provide the context needed to support change. Is the Fleet's documented performance with AIM-120 good enough to achieve success in the next conflict? Does that number need to be higher? Is a lower number still acceptable (i.e., capable of winning the next war)?

The governing military instructions that direct the Fleet Training Continuum and OFRP for the F/A-18 community outline a process for developing performance standards that start with the NCCs developing requirements to meet the needs of their specific areas of responsibility. Those NCC requirements are then used by United States Fleet Forces/Commander Pacific Fleet, the Type Wing Commanders (TYCOMS), and the Naval Aviation Warfighting Development Center (NAWDC) to develop the Navy Mission Essential Task List, Measures of Effectiveness (MOE) standards, and Measures of Performance (MOP) standards.

This process is sound in principle but falls short in execution. For the outlined process to be effective, it requires quantifiable data to effectively link the performance of individual skills to the desired outcomes of the NCCs. Without quantifiable data, any standards that were developed under the process remained largely subjective. More importantly, under the CBTRM system, which tracks warfighting readiness as a function of capability vice proficiency, performance standards for individual skills are not enforced. Finally, prior to the APM process, there was no mechanism available for collecting and measuring data to support the development or application of standards during training events. The APM process supplies the quantifiable data needed to inform the NAE's process. Additionally, the toolset being developed to automate that process (RAPID) also provides the means to transform that data into performance standards using AI/ML techniques.

Initial applications of AI/ML in the RAPID software use Multiple-Linear Regression Analysis (MLRA) for explanatory and predictive purposes. MLRA is particularly suitable for RAPID because the system being modeled (i.e., the combat performance of naval aviators) has structural features that are well understood and represented by established data metrics. Specifically, the APM data sets are based on previously defined metrics for aircrew skills (e.g., sensor mechanics, individual weapons employment, individual defensive execution, mutual support, etc.), aircraft conditions (i.e., system capabilities), and mission outcomes (fighter kills and fighter deaths). Furthermore, the relationship between the various input metrics and mission outcome is reasonably well understood by those with sufficient training and experience. Therefore, MLRA provides the capability for RAPID to: (1) explain outcome data in terms of multiple explanatory variables, (2) calculate an expected outcome for specified values of the explanatory variables, and (3) forecast future outcomes based on trends in, or hypothetical values for, the explanatory variables.

The first step in MLRA model development is to identify a response variable that represents the mission outcome. For F/A-18 air warfare, the mission outcome is best represented by the "kill ratio" (i.e., the ratio of fighter kills to fighter deaths). The next step is to identify a minimum set of explanatory variables that are highly correlated to the response variable and mutually orthogonal. Since the MLRA analysis is intended primarily to support training, each of these variables should represent a defined skill that could be a suitable subject of focused practice. Least squares regression is then used to estimate the model coefficients. Finally, the quality and reliability of the model is evaluated based on the effect size and significance of each explanatory variable, the R^2-adjusted statistic, and the confidence intervals.

The resulting MLRA model can be used to explain the relationship between kill ratio and the individual skills, to calculate an expected kill ratio for specified values of the skills, and to forecast future outcomes based on trends in, or hypothetical values for, the skills. The NAE has begun to rely upon this modeling approach in order to obtain the following benefits:

- Data-driven performance standards – By modeling the optimal set of skill values that are required to achieve a desired kill ratio, the NAE can provide clear objectives by which aircrew assess mission success in air warfare, linked to measurable outcomes in combat.
- Skill prioritization – By analyzing the effect that certain changes in skill values have on the resulting kill ratio, syllabi developers can identify which skills should be addressed first and most often.
- Predicted kill ratio – By understanding the likely range of kill ratios associated with a given level of proficiency in key skills, aircrew can focus their training efforts on the most important skills and improve mission planning (e.g., optimal aircrew/aircraft selection).
- Training benchmarks – By modeling the range of future kill ratios using demonstrated and extrapolated skill acquisition rates, the NAE can develop a roadmap for individual and unit performance across the OFRP (i.e., realistic training goals for each stage of training that culminates with the achievement of combat requirements at deployment).
- Periodicity requirements – By analyzing demonstrated and extrapolated skill decay rates, the NAE can establish data-driven periodicity requirements to ensure sustained performance.

The AI/ML logic supports the user in all the processes implemented in the software from mission planning through final debrief and analysis. For example, during pre-event mission planning, users enter details about the event as well as assigned aircrew/aircraft. These details are analyzed by the software's AI/ML algorithms in the mission planning process to provide the planner with expected outcomes for the event as well as recommended actions. The planner(s) can then explore potential changes to the event (e.g., in the assignments of aircrew, aircraft, and mission type) and the resulting impact on expected mission outcome or kill ratio. These features are provided to aid in aircrew selection, aircraft selection, aircrew/aircraft pairing, and game plan development. Once the final determination on aircrew is made, the software then prescribes ways to maximize the potential learning received during the upcoming event by developing a list of tailored training objectives. These tailored objectives ensure that each of the selected aircrew for the event is working towards improving skill deficiencies related to the selected mission and driven by their individual past performance. The software also prescribes ways to increase the likelihood of mission success by developing a list of "Keys to Success" (i.e., focus areas). These focus areas are tailored to the selected aircrew and are informed by skill deficiencies that have resulted in mission failure during past missions. By addressing these targeted deficiencies, the chances for mission success are increased.

After the event is flown, the users are provided an intuitive interface for importing data from the various disparate data sources. Once imported, the users are then logically walked through the APM process for reconstruction, analysis, and assessment of performance. Finally, the users are provided a summary of their individual performance and systems health measured against the standards linked to the desired mission outcomes, as outlined in the OPLANs by the NCCs. The AI/ML algorithms are also applied during this stage to provide recommended remediation steps to address any deficiencies identified during the event. These recommended remediation steps provide a realistic roadmap to improving performance to achieve the identified standards linked to mission success.

RESULTS

To date, TAP has collected data on six carrier air wings and two TOPGUN courses. The project has also documented performance of squadron-level exercises including detachments to St. Louis for virtual training. The current APM dataset consists of over 3 million data points and continues to grow. Data scientists at JHU/APL have conducted independent analysis of the APM data. They have concluded that the data is normal and were also able to validate, using a 95 percent confidence factor, the statistical significance of all categories of Fleet aircrew in the TAP data. As such, the TAP dataset is considered by the NAE to be statistically significant and actionable information.

Findings

TAP's first official data collection was conducted in 2015 during the execution of the Air Wing Fallon (AWF) syllabus. Naturally, this "snapshot in time" of a single carrier air wing did not produce a large enough sample size of data to generate any findings with statistical significance. It did, however, provide an indication of the importance of basic mechanics execution (i.e., skills associated with mutual support, sensor manipulation, communication, valid weapons employment, and correct defense execution) due to it being the single largest factor influencing mission outcomes. It also indicated that the lack of proficiency in basic mechanics was likely due to a lack of academic knowledge of the aircraft systems, the weapons, and the TTPs (i.e., a direct correlation between cognitive skill and resulting performance). Subsequent data collections, including observations of the Strike Fighter Advanced Readiness Program (SFARP) syllabus and TOPGUN course, have not only reinforced this finding with statistically significant results, but have also provided amplifying data on the training techniques (i.e., syllabi specifics, training environments, etc.) that have the greatest impact on skill acquisition/decay.

TAP data has also shown that air wing performance peaks during the early stages of OFRP training and tapers off during subsequent training evolutions that involve more complex scenarios. This is primarily because the focus of training over the course of an OFRP traditionally shifts from training basic mechanics early to the incorporation of integrated air wing operations as the unit gets closer to deployment. As this transition occurs, there is less emphasis upon basic mechanics (i.e., less instruction and less opportunities to maintain proficiency). This means that the early stages of OFRP training are critical and establish a foundation for the execution of individual basic mechanics that is critical to success later in the OFRP when operations become more complex. It also means that basic execution must continually be reinforced throughout OFRP training to maintain proficiency and ensure mission effectiveness. If basic mechanics are not mastered and maintained throughout OFRP training, air wings have little to no chance for success. Armed with this information, TAP led the effort to develop revisions to multiple OFRP syllabi including SFARP and AWF. The revised syllabi bring alignment to OFRP training, prioritize skills that have been proven to matter most in the context of desired NAE outcomes, and provide standards for skill execution across the OFRP that align to those desired NAE outcomes.

In addition to changing the way that the F/A-18 community trains, the APM data/analysis is also used to address the way that squadrons are resourced/equipped leading up to and during OFRP training evolutions. This is only possible because the APM data takes into account both aircrew reliability and systems reliability. The aircraft systems availability/health data collected by TAP is used to identify the importance of each individual system to mission success. This information, when paired with the training objectives established for each stage of the OFRP, results in a resourcing plan that ensures squadrons are equipped with the necessary systems at each stage of training.

The APM-derived syllabi and resourcing plan provide the NAE with a comprehensive strategy/roadmap for developing individual aircrew skills and systems capability across the entire OFRP cycle, culminating in performance that meets the needs of combat requirements at the start of their deployment.

Impact on NAE Training and Resourcing

Innovation is always accompanied with challenges, and TAP's implementation of the APM process across the NAE was no different. For years leading up to the development of TAP, the Naval Aviation community, much like many organizations both inside and outside the DoD, had been resistant to adopting data-driven processes. This was primarily due to the stigma that performance data would be used in punitive ways against the aircrew who were being observed. While data in the wrong hands can be dangerous, it is also required to make informed decisions. Without data, leadership must make decisions based on suppositions (i.e., personal bias) or estimates (i.e., guesses) regarding the overall capabilities and the training/resources that are required to win the next war.

The first hurdle that needed to be overcome when implementing the APM process was just getting access to Fleet aircrew to collect the data. During the early days of TAP, APM data collectors had to rely on personal relationships with their former peers to get explicit approval from air wing leadership to observe events. At the same time, the anonymity of the aircrew who were being observed had to be maintained at all times. This was done to ensure that the APM process, which was new and still developing credibility and validity, was not undermined by any potential or perceived negative repercussions that could come from reporting on individual and unit deficiencies.

The turning point for the effort was when NAWDC, the center of excellence for naval aviation training and tactics development, provided its full support for the APM data-driven performance assessment process and endorsed the APM results. NAWDC, which has a long tradition of developing ground-breaking approaches to training that are envied by other communities, recognized that the APM process provided capabilities that would allow Naval Aviation to adapt their processes, remain relevant, and achieve success in future conflicts. In 2018, NAWDC suspended all their existing data collection efforts, and now relies solely on TAP and the APM data collection process for their analytical needs. NAWDC's parent command, CNAF, now actively directs air wings to participate in the APM process (i.e., provides full support for the data collection effort and use of the resulting APM analysis) during their OFRP training. Likewise, the analysis has been expanded to include additional platforms and missions beyond just F/A-18 (e.g., E-2, EA-18G, F-35, etc.) and the NAE continues to make progress towards being data-driven for all training and resourcing decisions.

CONCLUSION AND FUTURE APPLICATIONS

The development and successful implementation of the APM process has been ground-breaking for the Naval Aviation community, which was initially resistant to documented performance assessment. The APM process is now a verified and validated methodology for analyzing individual, element, unit, and air wing level performance in a way that directly links to desired mission outcomes. The implementation of automated tools designed around the APM process are significantly increasing the quantity of data that is collected, enhancing the quality of the analysis, putting valuable information directly in the hands of the warfighter to improve combat capability, and enabling a transition from event-based training to performance-based training and resource management.

Although developed using F/A-18 air warfare as a test bed, the APM process was developed from the ground up to be platform, mission, and training environment agnostic. This means that the process is scalable and can be adapted to suit a wide range of analytical use cases, both in the DoD and commercially.

ACKNOWLEDGEMENTS

The authors would like to thank 2 Circle Consulting, especially the members of the APM team who implemented the process and were instrumental to its success. They would also like to acknowledge the contributions of JHU/APL for their technical support and data analysis. This research benefited from support from across the NAE to include NAVAIR, OPNAV, and CNAF. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Chief of Naval Operations (CNO). *A Design for Maintaining Maritime Superiority*. December 2018.

Vice Chief of Naval Operations (VCNO). *Perform to Plan (P2P) Governance*. 9 Apr 2019

Office of the Chief of Naval Operations (OPNAV). *3000.15A – Optimized Fleet Response Plan (OFRP)*. 10 Nov 2014.

Commander, U.S. Fleet Forces Command (COMUSFF); Commander, U.S. Pacific Fleet (COMPACFLT). *3000.15A – Optimized Fleet Response Plan (OFRP)*. 8 Dec 2014.

Commander, U.S. Fleet Forces Command (COMUSFF); Commander, U.S. Pacific Fleet (COMPACFLT). 3501.3D – *Fleet Training Continuum (FTC) Instruction*. 1 October 2012.

Commander, Naval Air Forces Pacific (CNAP). Commander, Naval Air Force Atlantic (CNAL) *3500.1 – Squadron Training and Readiness*. 10 Dec 2015.

Commander, Naval Air Forces Pacific (CNAP); Commander, Naval Air Force Atlantic (CNAL) *3502.1 – Air Combat Training Continuum (ACTC) Program*. 7 Apr 2016.

Commander, Naval Air Forces Pacific (CNAP). *5450.42 – NAWDC Missions, Functions and Tasks (MFT)*. 12 Jan 2016.

Commander, Strike Fighter Wing U. S. Pacific/Atlantic Fleet (CSFWP/L). *3500.7E – F/A-18 Combined Wing Training Manual*. 27 Jan 2014.

Commander, Strike Fighter Wing U. S. Pacific/Atlantic Fleet (CSFWP/L). *1525.1G – Strike Fighter Weapons and Tactics (SFWT)*. 4 Jan 2016.

Naval Aviation Warfighting Development Center (NAWDC). *3500.3H – Sierra Hotel Advanced Readiness Program (SFARP)*. 27 May 2016.