Establishing a Program for Applying Earned Value Metrics to Flight Test

Chris Rice
Robin Locksley

Naval Air Warfare Center Aircraft Division
22347 Cedar Point Road, Unit #6
Patuxent River, Maryland 20670-1161

Naval Air Systems Command
47123 Buse Road Unit IPT
Patuxent River, Maryland 20670-1547

Improving efficiency and decreasing costs are becoming more important as the pressure to procure aircraft weapons systems more quickly and at a lower cost increases. The current tool of choice for tracking program efficiency is Earned Value Management (EVM), which provides indices of cost and schedule performance against an agreed upon baseline for task completion. This paper discusses methods that will be used by the H-1 Upgrades Flight Test Team to implement an EVM scheme to track the efficiency of the flight test program. We will define EVM, discuss, compare the merits of existing metrics for flight test, and propose a database management approach. Bounding the problem and expected metrics are discussed. Finally we will present a methodology for uniformly planning for contingencies and unknown unknowns so as to permit success to be declared within the work package even in the face of technical challenges.

Earned Value Management (EVM), H-1 Upgrades Flight Test Team, EVM scheme

Unclassified

Unclassified

Unclassified

Unclassified

7

(301) 342-
Establishing a Program for Applying Earned Value Metrics to Flight Test

Christian B. Rice
Naval Air Warfare Center Aircraft Division
Test and Evaluation Engineering Group
Rotary Wing Mission Systems Division
22755 Saufley Road
Patuxent River, MD 20670
301-342-1969
doe@navair.navy.mil

Robin H. Locksley
Naval Air Warfare Center Aircraft Division
Test and Evaluation Engineering Group
Rotary Wing Mission Systems Division
22755 Saufley Road
Patuxent River, MD 20670
301-757-0605
locksleyrh@navair.navy.mil

Abstract—Improving efficiency and decreasing costs are becoming more important as the pressure to procure aircraft weapons systems more quickly and at a lower cost increases. The current tool of choice for tracking program efficiency is Earned Value Management (EVM), which provides indices of cost and schedule performance against an agreed upon baseline for task completion. This paper discusses methods that will be used by the H-1 Upgrades Flight Test Team to implement an EVM scheme to track the efficiency of the flight test program. We will define EVM, discuss, compare the merits of existing metrics for flight test, and propose a database management approach. Bounding the problem and expected metrics are discussed. Finally we will present a methodology for uniformly planning for contingencies and unknown unknowns so as to permit success to be declared within the work package even in the face of technical challenges.

TABLE OF CONTENTS

1. INTRODUCTION
2. TRADITIONAL EARNED VALUE METHODS
3. TEST POINT DATABASE
4. DEVELOPING WORK PACKAGES
5. ACCOUNTING FOR UNKNOWN UNKNOWNS
6. DATA COLLECTION PROCESS AND TAKING CREDIT
7. COST AND BENEFITS
8. PERFORMANCE METRICS
9. CONCLUSIONS AND RECOMMENDATIONS

INTRODUCTION

One of the requirements of the H-1 Upgrades Program is to use Earned Value Metrics (EVM) throughout the program; including flight test. Historically, the use of EVM in flight test has had limited value although more attention is bringing about positive change. Traditionally flight test progress (for an envelope expansion test program) is presented on a V-N (velocity versus normal acceleration) diagram. Figure 1 shows a typical V-N diagram for a helicopter with the edges of the design envelope shown.

When the envelope is initially expanded, the concentration of data points is around the hover point. As confidence in the aircraft increases, level flight sweeps are conducted. Following level flight sweeps, a middle of the envelope airspeed will be chosen to examine increasing load factors. This presentation while valuable does little to support the management process. Where does the leader need to focus? What are the inhibitors to progress? Are the parts supply and repair systems supporting the test program properly? These types of questions need to be answered to support a hard charging, efficient flight test program. Our hope is that earned value management will provide the necessary management insight to easily answer some of these questions.

*Earned Value Management (EVM) is the use of an integrated management system that coordinates work scope, schedule, and cost goals and objectively measures progress.*

---

1 U.S. Government work not protected by U.S copyright.
2 As taken from [1].
toward these goals." [2] By using EVM, the program manager can determine what areas of their acquisition program are performing to cost and schedule. This traditional use of the EVM is generally defined as a contractual requirement and is managed by the contractor, who report the results to the government acquisition team. Our desire is to develop metrics to provide insight to the performance of the aircraft against the design criteria as well. For example, is the hover performance of the aircraft progressing in such a manner that program managers can expect to meet the operational requirements? This kind of information, will be invaluable to decision makers if the need to reprogram program funding arises. In addition to providing the program management team with overall metrics, we hope to provide metrics that will improve the efficiency and effectiveness of flight testing along with data for lessons learned to support future test programs. One of the benefits that should become obvious is the improvements in the efficiency and effectiveness of the flight test program that result from the detailed planning necessary for the development of the EVM techniques.

In this paper, we will discuss one proposed application of EVM to the H-1 Upgrades flight test program. We will present a brief overview of the earned value management process. We will then present the requirements of a test point database, discuss the importance of early definition of all test points, the steps we have taken to ensure commonality amongst test data collected, and the necessity to account for planned non-test activities. Then, having laid the framework from which to work, we will discuss our plans for data collection and disposition, how to manage the program baseline, and the various metrics that will be available from the data collected. Finally, we will conclude and present recommendations for future work.

2. TRADITIONAL EARNED VALUE METHODS

EVM is a method of systematically measuring progress toward the goal of project completion. Simply measuring money spent and work completed as measures of progress does not sufficiently answer the question. Is the project on track? The following EVM terms are defined [3]:

BCWS Budgeted Cost for Work Scheduled. The project baseline or plan that encompasses all work packages, funding, and scheduling.

BCWP Budgeted Cost for Work Performed. The actual performance or earned value of the work performed.

ACWP Actual Cost of Work Performed. The actual project cost as opposed to the projected cost.

BAC Budget at Completion. The total cost of all of the budgeted plans.

EAC Estimate at Completion. The forecast of the actual total cost at completion based on the BCWS and the ACWP to date.

Figure 2 shows a cumulative plan/status display for a typical project. In this presentation, project status in terms of cost and schedule is clearly seen. In this example, the project is behind schedule (BCWP lower than BCWS) and over budget (ACWP higher than BCWP).

![EVM Cumulative Plan/Status Display](image)

Figure 2. EVM Cumulative Plan/Status Display

It is well documented that the trends in schedule and cost are established as an indicator for the entire program, if the project is more than 15% complete [2]. After this point budget overruns cannot generally be recovered and schedule slippage tends to grow rather than shrink. The two key performance indices that may be calculated from data presented in the EVM format are cost and schedule efficiencies or cost performance index (CPI) and schedule performance index (SPI):

\[
\text{Cost Efficiency} \quad \text{CPI} = \frac{\text{BCWP}}{\text{ACWP}} \quad (1)
\]

\[
\text{Schedule Efficiency} \quad \text{SPI} = \frac{\text{BCWP}}{\text{BCWS}} \quad (2)
\]

Where CPI or SPI > 1.0 is favorable and < 1.0 is unfavorable.

These indices are aggressively tracked by senior managers within the DOD. In the current budget environment, if CPI or SPI are unfavorable at early milestones in a program, it is at risk of cancellation or additional oversight.

As taken from [3].
3. TEST POINT DATABASE

One of the first issues to deal with is to define what is needed to provide the measures of effectiveness, what do we need to know to develop the EVM data. Based on the experience of other test programs, it is critical to design a database to manage the test data. Databases exist that have been used by test programs and we are attempting to draw the best from each.

Bounding the Problem

In order to effectively manage the development and envelope expansion phase of a flight test program it is necessary to programmatically define the areas of the flight envelope that will be initially examined. In the H-1 Upgrades program, the test and evaluation (T&E) leadership limited the test team to the following conditions per type/model/series:

a. 3 Gross Weights
b. 3 Centers of Gravity
c. 4 Altitudes
d. 8 Configurations of Ordnance

These constraints forced the engineering teams to think in terms of heavy, medium or light gross weights; forward, medium, and aft CG configurations, hover, low, medium and high altitudes; and the eight fixed ordnance load-outs when formulating their test plans. This train of thought built in the flexibility necessary to allow the test team leadership to investigate where concurrent data collection could be used. While there are some conditions that will not fit into these groupings, these are handled on a case by case basis in order to minimize their occurrence and verify their validity.

The Database

The test point database is the soul of the machine used for measuring the progress of the test program. For the H-1 Upgrades program, the Microsoft Office™ family of products was chosen for data manipulation. The database was constructed using Access™ with the test matrices drafted in Excel™. Excel™ was chosen for the test matrices since most test engineers are conversant in Excel™ but have little experience in relational database products such as Access™. These programs easily permit data to be exchanged between the two.

If the database is developed properly, many data products may be derived from it, including:

a. Test team performance indices
b. Progress towards test plan completion by period, aircraft, or test team
c. Sortie data cards
d. Program performance summaries
e. Graphics of event completion

Using lessons learned and databases developed for other flight test programs as a starting point, a database is under construction that contains all the test points deemed necessary during the planning stages of the flight test program. It is critical that all data points be defined by similar initial conditions:

a. Maneuver type index (MTI)
b. Loading
c. Category of test
d. Main rotor RPM (Nr)
e. Automatic Flight Control System
f. Configuration
g. Altitude
h. Airspeed
i. Gross weight.

Achieving this goal, the test plan format was supplied to the engineering teams with pre-defined headers in the test matrices. Additionally, the test matrices contained additional fields to capture remarks, methods, time required to be "on condition" or other pertinent data or conditions required. When the database is fully populated, it will be sorted by initial condition to identify data points with similar requirements. This will allow the team to plan and manage data collection more effectively by maximizing the number of data points collected at each point in the sky. Reducing the number of times the aircraft has to be at a specific condition of airspeed, altitude, gross weight, etcetera, will naturally improve efficiency.

Prerequisite or build-up test points are also included in the test point database in addition to the end points. If testing goes well and larger increments can be safely taken, credit can be obtained for the intermediate points that were omitted, without penalty. The catalog of initial conditions should also support the specific conditions for all qualitative test points. However, the test team is still evaluating the best method for tying qualitative test conditions to specific initial conditions.

From a programmatic perspective, all flight test related events must be accounted for through the course of the test program, such as:

a. Planned maintenance events (ground functional test and flight events like functional check flights)
b. Aircraft modification periods
c. Ground tests
d. Pilot training
e. Guest and public relations flights
f. Ferry flights
g. Chase flights
h. Familiarization flights

This additional level of detail, while not test point related, is critical because the database is what feeds the flight test performance metrics and is used to assess program progress.
4. DEVELOPING WORK PACKAGES

A work package is a measurable quantity of work within the EVM system. Traditional EVM techniques strive to create work packages at the lowest possible level in order to make the package as distinct and short as possible. If this is done properly, work packages should be finite enough to illustrate progress, yet detailed enough to provide management insight into problems that may occur. For longer duration work packages, the package should be subdivided into discrete value milestones that allow objective measurement of work performed. Within the flight test program the work packages can be defined by test area, and the work packages will include roll ups of test conditions in the database. Using handling qualities as an example, several work packages can be broken out; longitudinal, lateral-directional, and low airspeed performance. If the work packages are based on sorties or data points, they become too granular; progress will show up as individual points on the V-N diagram. Conversely, using handling qualities as a work package would provide little insight into how testing that aspect of aircraft performance was progressing; the scope is too broad. As the test program evolves, some data points will need to be re-flown and the work packages should be developed to accommodate a reasonable amount of retest.

Weighting the Work Package

Based on lessons learned, consistency between the value of the work packages should be considered or the EVM credit will have peaks in the cost and schedule performance curves. Hence, the hover performance work package and the functional check flight work package must carry the same weight in the overall program's EVM system. This is based on the fact that all work packages will require labor hours and schedule regardless of the nature of the work being performed, or whether they provide progress towards the next test event or to ensure aircraft readiness. If the guest flights for the month are less than planned, credit should be taken, because there is a good chance that the next period may have more guest flights than originally planned.

Systematically estimating work package duration

The next step in developing the work packages is to accurately estimate the calendar time required obtain and analyze the test data. A roll-up of all of the work packages will establish the number of test months required to complete the program. The Navy's V-22 Integrated Test Team [4] used the following equation to systematically establish the rate at which data could be collected:

\[
\text{Productivity} = \frac{\text{Events}}{\text{Month}} \times \frac{\text{Sorties}}{\text{Month}} \times \frac{\text{Test Rate}}{\text{Hour}} \times \frac{\text{Data Rate}}{\text{Hour}}
\]

Where:

- **Productivity**
  - Number of data points that can be collected per month.

- **Sorties**
  - Productive flight hour rate that can be generated

accounting for maintenance, unforeseen maintenance issues, weather, and configuration changes.

**Test Rate**

Actual percentage of flight during which testing can be conducted. Accounts for start/taxi/takeoff and transit to ranges.

**Data Collection Rate**

Rate at which data points can be collected accounting for automated instrumentation, pilot workload, and condition set-up time.

The rate at which sorties can be generated varies from 10 flight hours per month for complex fly-by-wire aircraft, to 30 flight hours per month for avionics black box testing. Test rates generally average 75-90%. Prior planning can increase this percentage by providing the aircrew data cards for tests that can be conducted while in transit to and from the operating area. Data collection rates vary from 10-120 data points per hour, depending on the complexity of set-up or duration the maneuver must be sustained. Figure 2 shows a comparison of the flight hours planned for collecting data by test type for the H-1 Upgrades program.

**Modification Periods and Guest Flights**

It is important to consider all the possible events that are part of the normal course of a flight test program when developing work packages and this data needs to be included into the database of test points. Most test program will have modification periods within the program for instrumentation changes and equipment installations. All programs also have the benefit of high visibility guest pilots that should be accounted for in the schedule with equal event weight. During the planning stages, estimates of schedule and cost impacts of these periods should be included in the EVM plan. If the events occur as scheduled, the EVM data will reflect that. If the events take more time than estimated, the EVM data will reflect that via a negative trend. If the modification period is not required, credit for completing the period should be taken [4]. If the planned modification period is not required, test program efficiencies must have precluded the need and EVM credit will reflect that. Without this careful planning up front, these events will result in falling behind schedule and exceed cost through no fault of the test team.

5. ACCOUNTING FOR UNKNOWN UNKNOWNS

By its very nature, flight testing is designed to uncover problems. An analysis of past programs will provide areas where additional data points have been historically necessary. The prudent test team will add additional data points in these areas to address these potential contingencies up front. However, what about the so-called "unknown" unknowns? Solutions to these problems are often iterative, and require indeterminate resources for their resolution. Since they are unforeseen, planning for them is impossible. In these situations, cost and schedule performance indices will show unfavorable results that do not indicate flight test inefficiency but do reflect progress towards test program
Cooling and Temp Survey
Electrical Systems Demonstration
H-V & Performance
Handling Qualities
Power Plant Demo
Aircraft Ferry
Weapons Accuracy
Firing Loads and Vibes
LowA/S System Calibration
Target Sight Development
Avionics Validation & Weaps Checks
Avionics Development
Structure Demo
Non-Firing Vibes
Non-Firing Loads
Envelope Expansion
Development Tests

Figure 2. Predicted Flight Hours by Test Type (AH-1Z)

completion. At this juncture, flight test efficiency and program efficiency diverge.

When problems are uncovered in the course of flight testing that exceed the planned work the CPI is negatively affected in attempts to maintain schedule. From an overall program perspective, these negative trends in CPI and potentially SPI must be observed to correctly budget for completion. However, the identification of a problem may or may not affect flight test efficiency. For example, if a problem requires additional sorties to investigate, and fixes can be implemented within the period between planned sorties, then the sortie rate (a measure of flight test efficiency) will not be negatively affected. Conversely, if the investigation and resolution of a problem requires grounding the fleet, then flight test efficiency will quickly be affected through no fault of its own.

From a programmatic perspective, unknown unknowns should be accounted for within the overall contract effort, but since flight usually uncovers unknown unknowns, the tendency is to credit flight test with those EVM impacts. To attempt to estimate the flight test cost and schedule impact of the unknown unknowns becomes a real challenge. For example, assume that during the envelope expansion testing, a tail wag problem is discovered. To address a tail wag problem will require engineering, manufacturing, and flight test hours. Once the design engineers have sufficient data from flight test, a new approach evaluated. Manufacturing personnel must produce the necessary parts and an aircraft modification period is required to make the configuration change. Once the modifications are completed, the new configuration must be tested. This process might be repeated several times before the tail wag characteristics of the aircraft as sufficient to meet mission requirements. The flight test time to evaluate the new configuration could be significant and include second looks a the previously cleared portions of the flight envelope.

The labor hours to correct the problem should be charged to design engineering vice flight test with respect to the EVM data. Schedule will also be required to correct the problem. Getting the schedule and cost issues addressed in the proper labor areas will be a challenge for the flight testers. The schedule impacts will have to be absorbed into the existing test schedule since the end point of developmental testing usually fixed by an acquisition milestone. Some additional schedule might be available to the developmental testers if the operational test community is intimately involved throughout the test program.

The obvious question becomes: “How do we address the unknown unknowns in the EVM program?” In discussions with other programs, the common approach is to pad the number of events or pad the event budget. In this program, we will likely pad the event budget; however, because the test schedule is fixed, padding the schedule budget will not be possible.

6. DATA COLLECTION PROCESS AND TAKING CREDIT

As the test program begins, the database of required test points becomes the focus for documenting test completion. Of particular importance is how the engineering team dispositions data points so that credit (BCWP) can be taken for their execution. The F/A-18E/F Integrated Test Team used the following categories for dispositioning data [5]:

TESTED
The data point was successfully flown and data processing was pending.
BAD PT (Bad point)
Attempt to collect data that was unsuccessful due to test conditions.

BAD DATA
Attempt to collect data that was unsuccessful due to test failed critical parameters.

NDR (No Data Required)
Real-time telemetry of pilot confirmation sufficient to consider point complete.

After an appropriate waiting period for the data to be reviewed, the test point’s disposition was changed to DATA (indicating that the TESTED data had been reviewed and accepted) NDR, or CANCEL (which was reserved for test points determined to be impractical or not achievable). Requiring data points to be dispositioned promptly will permit timely metrics to be extracted from the database. BAD PT and BAD DATA required data point to be re-flown. This effort was considered unplanned work.

7. COST AND BENEFITS

In keeping with current Department of Defense acquisition policies, contracts require the use of EVM throughout the program. As such, the cost of using EVM for flight testing has been factored into in the contractor proposal. However, we in flight test are striving to collect additional data, which will not be free. The primary cost of establishing the EVM system is in the planning stages. Once the flight test program begins the data collection and reporting will be automated to the maximum extent possible. This additional cost of preparing the database to identify and manipulate the EVM data is low if it is completed early in the test program planning. Some additional cost may be to training the test team to make inputs to the database or to modify the database if a re-baseline is necessary. We do not expect these cost to exceed a half work-year in labor.

The benefits of using EVM in the flight test program should be significant. From a historical perspective, future test programs will be able to evaluate this test programs for lessons learned and have the data to present to the program managers in a format they are familiar with. For example, how many flight hours per month can be expected during weapons testing? Improvements in test planning for future program should be possible in many areas;

a. Flight Hour Estimates
b. Guest Flights
c. Modification Periods
d. Test Efficiency/Effectiveness
e. Weather Days
f. Planning

Additional statistical data should be available to evaluate if the size and composition of the test team and if the scope of the test program is compatible with the proposed schedule.

During the current test program, aggressive use of EVM data should also provide significant benefit to the test team and program managers. The test team leaders should be able to use the data as an indicator of the health of the test program and areas that are in need of management focus. As discussed earlier, unexpected events will occur during flight test and EVM data should provide insight to management when reprogramming test events becomes necessary. In other words, management attention can be directed when it is needed, as indicated by the metrics.

8. PERFORMANCE METRICS

The results of the EVM program will be the performance metrics. The question to ask is what we do want to get from our metrics, what do we want to measure? The information to be provided to the program manager is known and defined by DOD acquisition guidelines - cost and schedule performance. The flight test team will reap benefits if the proper measures are defined. One thing is for certain, unless significant events dictate otherwise, the flight test team is very unlikely to exceed the original schedule since a program acquisition milestone generally drives the end date.

Of primary concern to the developmental tester is will the aircraft successfully complete the operational evaluation? Progress towards meeting the design performance parameters can be measured against the plan. Is progress toward meeting the prerequisite events on schedule, are labor hours tracking to the planned events? If the first modification period can not be completed on schedule, it should provide insight to the remaining scheduled modification periods. Some additional metrics are shown below:

a. Is enough time allotted for test evolution?
b. Are the maintenance delays and inspections estimated accurately?
c. What were the maintenance delays?
d. Is the maintenance less intensive that anticipated?
e. Is the use of military maintainers effective in preparation for operational test, or was test productivity impacted?
f. Is the parts and spares plan working well?
g. Is the use of simulation effective?
h. Were there specific times of the year that showed an improved productivity?
i. Were the design issues discovered during test predictable?
j. Why were tests aborted? Were they ground or in-flight aborts?

With these measures available, the test team should be able to use metrics to constantly adjust personnel and the remaining schedule to improve the chances of completing the program as defined by the program manager as well as providing lessons learned for future programs.

9. CONCLUSIONS AND RECOMMENDATIONS

One of the more obvious conclusions is that the use of EVM requires significant up-front planning to be successful.
Estimates of all known test and non-test related events that can impact the test program must be considered, from maintenance concerns to flying VIPs and logistics issues. Also addressing the unknown unknowns within the baseline is critical because, flight testing is designed to find problems and in spite of the best efforts of the designers, unknown events occur. The additional approach of bounding the test matrix should provide interesting results.

The value to EVM to flight testing is significant. It is a language that program managers understand and provides real insight to the progress of the test program. If planned properly, the test team should be able to test more efficiently and address problem areas more effectively, providing more bang for the flight test buck.

We will continue this approach for use by the H-1 Upgrades test team and work within the flight test community to learn more about how test team are tracking progress and what metrics are used. As more data becomes available from flight test programs in the form of lessons learned, more efficient test programs should result. An evaluation of the results of this approach to EVM is the next step.

REFERENCES


Mr. Robin Locksley is a civilian flight test engineer working for the US Navy. He has a BS in Electrical Engineering from Drexel University and an MS in Electrical Engineering from the Florida Institute of Technology. He has worked as a rotary wing mission systems flight test engineer for 10 years, and is a graduate of the US Naval Test Pilot School. He gained much of his knowledge of flight testing while working as a project engineer for avionics and mission systems on H-3 and H-60 series helicopters, and as the lead engineer responsible for the testing of integrated ship/air elements of the SH-60B LAMPS MK III weapons system. He is currently the Flight Test Integrated Product Team Leader for the H-1 Upgrades Program.

Mr. Chris Rice is a civilian flight test engineer working for the US Navy. He has a BS in Aerospace and Ocean Engineering from Virginia Tech and an MS in Aviation Systems from the University of Tennessee. He has worked as a rotary wing air vehicle and mission systems flight test engineer for 15 years, and is a graduate of the US Naval Test Pilot School. He gained much of his knowledge of flight testing while serving as project leader for avionics and mission systems flight testing of the US Marine Corps MV-22B and during flight testing of US Navy H-60 series helicopters. He is currently a competency manager in the flight test of installed helicopter mission systems hardware and software.