Better Metrics for Better Communication as a Foundation for Better Program Buying Power

Dialogue Between Program Oversight and Program Execution Organizations

Brian Brodfuehrer

In his Better Buying Power memo, Under Secretary of Defense Dr. Ashton Carter recommended several actions, two of which were: "Reduce the number of OSD-level reviews to those necessary to uncover and respond to significant program issues" and "Reduce non-value-added overhead imposed on industry."

This intrigued me, because both actions were related to data I had recently gathered to facilitate a discussion dealing with the "program execution versus program oversight" nature of metrics, especially as related to cost estimates.

To support that discussion, I interviewed people with knowledge of and experience with the metrics good program managers use. They represented many years of acquisition experience in different product domains and stakeholder perspectives. For this article, I have quoted them anonymously.

Developing an effective set of metrics to be used by two organizations with differing interests establishes the context for this article's development, a shared emphasis on cost focuses that context a bit more.

Brodfuehrer *is a faculty member in DAU's PMT 401 course. He has over 30 years of acquisition experience working for both the government and industry.*

Metrics: The Good, the Bad, the Ugly

One thing that makes metrics ugly is that they are hard. It takes a lot of work to establish metrics properly, gather data on a regular basis, analyze the data, and decide how to act on it. Several experts commented on this: "If the value gained is less than the cost of gathering the metrics, don't do them." In other words, "You have to determine if the juice is worth the squeeze."

Metrics are also ugly because they can result in people being treated poorly. One senior acquisition program manager with experience in DoD and non-DoD systems said, "All federal acquisition is not high-trust. Honesty often results in axing heads."

An Air Force program manager working on a major joint aircraft program put it this way: "Metrics should measure processes, not people. The culture needs to be set [such] that when a metric goes south, management will seek to correct the processes (tools, training, and resources) and not take action against the people."

Another ugly aspect of program management metrics is dealing with numerous stakeholders and a rapidly growing IT environment. A large number of stakeholders represent distinct and varied interests, and each often desires different and more data. That desire is reinforced by the growth of information technologies that promise to gather and distribute larger amounts of data, faster, and to more people. Such growth sets high expectations for the art of the possible. Both trends push for more metrics, when fewer may be better. The ease of getting data can lure managers from focusing on the value proposition that the metrics are supposed to improve.

One source summed it up: "Knowledgeable stakeholders should pick a few insightful metrics and motivate the entire team to respond to them is the way to go rather than to gather a lot of data and not do anything with it."

Next, the "bad" of metrics: Program management is often about dealing with turbulence and bad turns of events. The program management environment, especially on

major programs, is one of responding to constant "I need data now!" fire drills. Frequently, data can portend bad news, and bad news does not get better with age. Not only are these drills time consuming, but they also take the PM's focus off program execution to attend to program explanation. A well understood, timely, shared, and consistent set of metrics can enable navigation in this bad environment.

In short, a good metrics program can soothe the savage oversight beast, meeting its need for information and tending to its fear of surprises. This is likely a harsh characterization of some oversight functions, but it represents a view from the execution side.

Another bad aspect of metrics has to do with the observer effect. Whether in physics or in the social sciences, the very act of taking a measurement can affect what is being measured. This can encourage bad behaviors leading to attitudes embodied in words like "spin." Care must be taken to avoid creating bad side behaviors when developing a metrics program.

"I can recall when I was a captain (Air Force), and I would be in the plant and spend long hours in the evening watching a software test or a qualification test and knowing exactly what happened. The next morning, I would sit in on the management team meeting and listen to their metrics for the event. I wondered if we were talking about the same event."

Finally, the good: A good side of metrics deals with empowering people by answering questions like "Where are we headed?" and "How are we meeting our value proposition?" When metrics show team members how they are achieving the program's objectives, they are more willing to set challenging goals and work hard to meet them. This is especially so when management or other partners, using the same metrics, provide the resources needed to meet the warfighter's needs while leaving more money in the taxpayers' pockets. "Think of it this way: Metrics don't only measure behavior; they drive behavior," said one program manager.

Another good use of metrics is to enable the use of relevant and timely data instead of conjecture to make decisions. Good metrics encourage the dual, technical-social nature of program management; their data is solid and they motivate productive human behavior. One more good side of metrics is that much work has been done to develop a large number of programmatic metrics and to organize them in a way that they can be used effectively across the enterprise. This article assumes the reader is aware of methods such as probability of program success (PoPs) in use across the DoD enterprise.

Metrics and Cost Growth: Differing Estimates, Errors, Decisions and Execution

One thing that can shape the discussion between program execution and program oversight organizations is when they have differing cost estimates. These represent both challenges and opportunities.

A challenge is how to determine which estimate is "right" and thus which number to use for budget purposes. Another is how to work the people side of coming to an agreement.

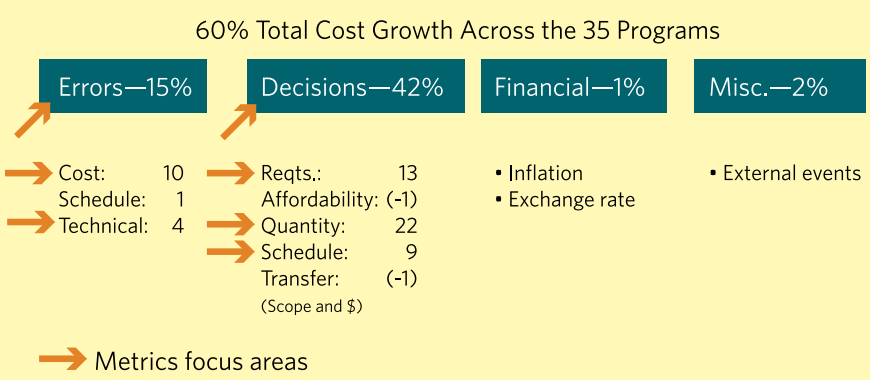
Opportunities exist, too. Identify the differences between the estimates, and look at the assumptions that drive the differences. These differences represent prime areas where metrics could be shared between organizations. Assumptions could be tracked over time to determine which ones materialized. Where are the unknowns in the two estimates and how will they be clarified? Unknowns drive cost. Differences in estimates represent areas where more dialogue may be needed to better have a common understanding of the program. Where estimates are the same there are opportunities too. The estimate could be right on, or perhaps it is too conservative, and cost could be reduced.

Working through the process of using the data from differing cost estimates to select a mutually agreed to set of metrics can: improve long-term communication, enable better joint decision making and reduce the oversight burden—all of which will have a positive impact on program affordability.

Errors and Decisions

In 2008, a RAND Corp. study showed that total cost growth for 35 major defense acquisition programs was dominated first by decisions made after the baseline estimate and second by errors in the baseline estimate. Decisions accounted for more than two thirds of the growth and errors for a quarter of the growth. Thus decisions and errors, in that order of priority, could be a useful way to further focus the dialog between the organizations. (See Figure 1.)

Figure 1. RAND Study: Sources of Cost Growth



Let's take errors first. One area impacting errors is proper estimation of the amount of design work versus true commercial-off-the-shelf (COTS) work in the program. Work requiring some type of design such as modified COTS differs from non-developmental item, re-use, or heritage. Metrics tracking these areas can be used to show trends as the program matures against the original estimates.

"Metrics should measure processes, not people. The culture needs to be set [such] that when a metric goes south, management will seek to correct the processes (tools, training, and resources) and not take action against the people."

Said one PM: "Program managers need to know how much of their program is or is planned to be re-use. This goes for both hardware and software. Will it really be COTS, or will it be modified COTS? For a program that is using existing hardware, do you know how many obsolete parts are in the boxes? How about diminishing manufacturing suppliers, how many of these will your planned effort have to deal with? Tracking and understanding these types of metrics can change your whole management approach. You may have signed up for a production program and find out that in reality you have one very much in development."

This concept can also be used in a related way for software. "Track how many software modules need to be designed. If the number is increasing then you know you have a growth problem. It could be due to adding more to the program or to not understanding the original task."

Another cost growth driver from errors in estimates are the technical issues the program faces, or, its technology maturity. Here, technology readiness levels (TRLs), when properly applied, are a helpful measure. Progress in the development of the technology along the TRL continuum, as compared with what was planned, can be tracked as a metric.

Known Unknowns

A third element affecting cost growth is capturing and dealing with the numbers of unknowns programs typically have early in their life cycle. A list of unknowns and decisions that need to be made across the acquisition spectrum (cost, schedule, performance, risk, stakeholders) can be made and progress tracked to completing actions necessary to bring clarity to the unknowns. Dealing with these unknowns will affect cost growth due both to errors and decisions.

Next, let's deal with the decision aspect. The RAND study pointed out that decisions made by the government after the original estimate to change quantities, add requirements or change the schedule significantly impact cost growth. Metrics after Milestone B will be necessary to ensure that decisions made are not causing requirements creep.

Examples of these might be: how many inter-agency memorandums of agreement are required and how many are completed? How many interface control drawings (ICDs) are required and how many completed? How many unknowns, such as "to be determineds" (TBDs) are in the ICD or specification documents? And what is the plan to burn them off?

"Most programs track the number of Class I changes. But what about the Class II changes? You need to have a way to ensure that a series of Class II changes won't add up to bite you."

"How are you identifying your unknowns and trying to put certainty into that uncertainty? What are the key decisions and when do they need to be made to keep the acquisition on track?"

"When doing software testing, you can't test every possible state. Also some coders leave paths behind that are terminated and not to be used. How many of the total population of possible states you could get into have you tested?"

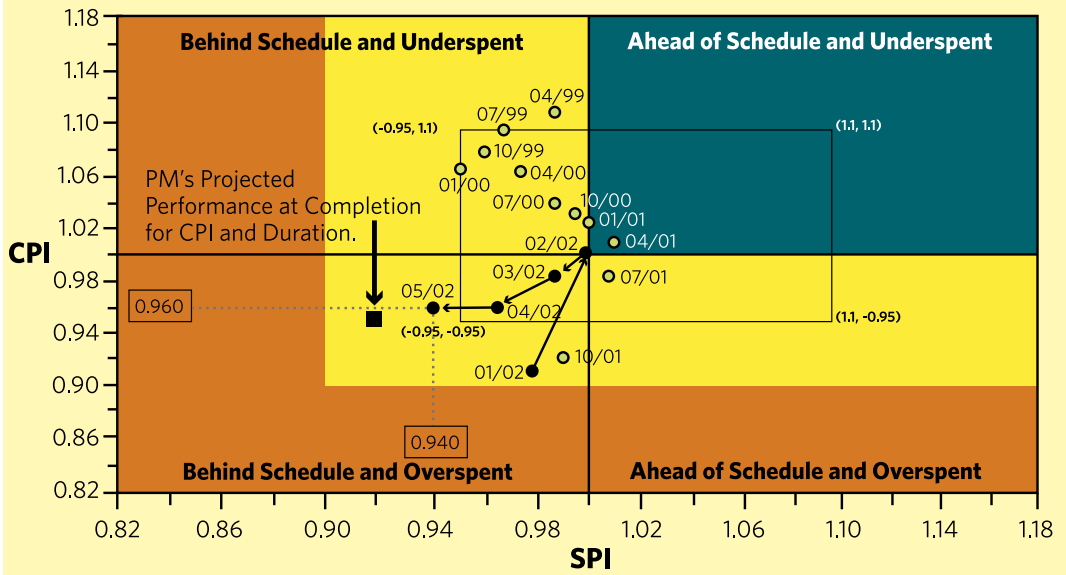
"How are you managing risk? The risk management program in itself is a series of metrics and success in controlling cost growth, especially in the technical maturity area, is directly related to managing the risks you have."

Using metrics to bound errors from the original estimates and to monitor and control decisions made after program start can improve program affordability.

Execution-Based Metrics

The ultimate goal of any set of metrics is the ability to forecast the future with enough lead time for actions which can effectively improve the predicted outcome. A challenge with metrics is that they are only as good as the assumptions made when they were created. Frequently as time passes, while the metrics collected are dutifully plotted and tracked, the assumptions behind them are forgotten rendering the current PM to only guess at their real meaning and usefulness.

Figure 2. Execution—Contract Earned Value Metrics



"EVM is a great tool, but the assumptions are important. How is the contractor measuring progress on a work package? Does he take 75 percent just for opening the package? That is a problem."

"One of the best ways I have seen to use EV data is to plot out SPI and CPI for each month and look for trends. I would draw a box around an area of small change and within that box consider the variations normal program jitter. What I looked for was trends, it worked great. And it was also useful to look at the program after a re-baselining; [seeing] the same trend means the original problem was not fixed." (See Figure 2.)

"A complete integrated master schedule (IMS) is important. I once asked to see one of my new program manager's IMS and what they showed me just represented the contracted part of the effort; it did not include other partners' or the government's part. It is not an IMS without the whole picture."

"A good IMS is critical; everything else depends on that. The earned value, everything. How many orphans are in the schedule? If a task has no parent, then why is it part of the plan? If it has no children, then how big is it, and how much work is required? Those represent unknowns."

A final challenge with execution metrics is that frequently the acceptability of the message is more dependent on the messenger than the data provided. Several of the managers interviewed said they would watch for this constantly, because a skilled messenger can make even sour milk taste sweet.

"I liked to pull a cost account manager's (CAM) name from a hat and ask that person to brief their status rather than rely only on the company to choose the CAMs who present at meetings."

"Industry and the government should use the same set of books. There should be almost real time access to data by the government after it is posted by the contractor."

Good execution-focused metrics enable better decisions and better decisions improve program affordability.

Summary

Carter's memo challenges the acquisition community to vigorously find ways to improve the buying power of the Department so that it can better face a challenging threat in a climate of constrained budgets. This will require, once again, a change in culture or mind set on the part of the acquisition community. The last quote, about the power of culture to navigate conventional acquisition systems to deliver capability to warfighters, comes from a program manager who worked acquisition for the Special Operations Forces (SOF) community:

> There is a myth that the Special Operations community executes acquisition through unlimited funding, higher priority, dodging the rules. The reality is that yes, there are benefits to having a smaller community and a more direct line to decision and budget authorities, but these benefits are offset by more demanding operational environments and higher customer expectations. The difference is the way SOF gets it done. They have a 'can do, must do' attitude that enables them to navigate through conventional acquisition systems to deliver capability.

Metrics drive behavior, and taking the time to establish those better metrics can create the attitude and communication necessary to satisfy the varied interests of stakeholders and improve the buying power the whole acquisition community in years ahead.

Finally, productive dialogue between execution and oversight organizations will be central to the way forward and hopefully this article has provided ideas, based on actual practice, to guide that dialogue.

The author can be reached at **brian.brodfuehrer@dau.mil.**