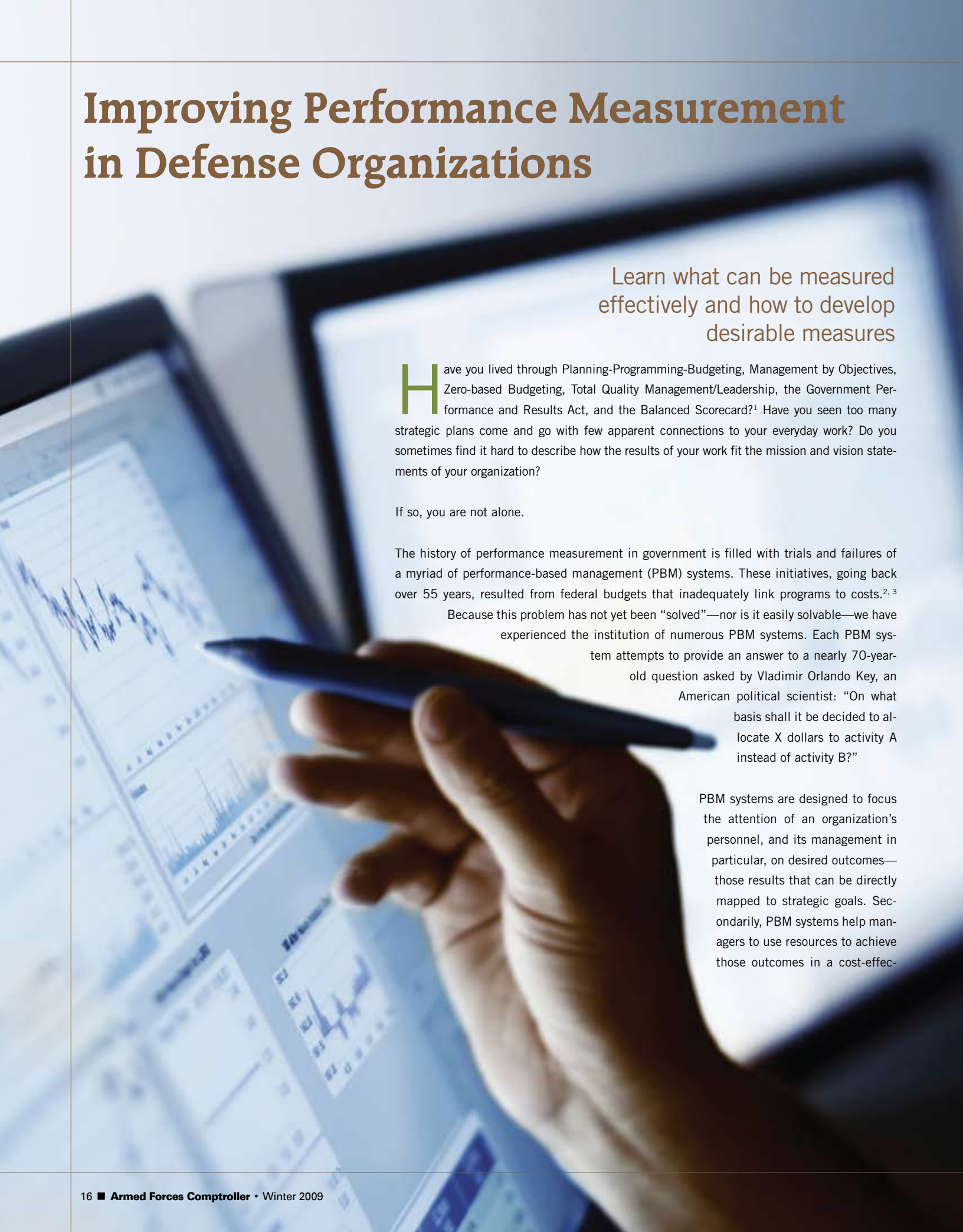
Improving Performance Measurement in Defense Organizations



Learn what can be measured effectively and how to develop desirable measures

Have you lived through Planning-Programming-Budgeting, Management by Objectives, Zero-based Budgeting, Total Quality Management/Leadership, the Government Performance and Results Act, and the Balanced Scorecard?[1] Have you seen too many strategic plans come and go with few apparent connections to your everyday work? Do you sometimes find it hard to describe how the results of your work fit the mission and vision statements of your organization?

If so, you are not alone.

The history of performance measurement in government is filled with trials and failures of a myriad of performance-based management (PBM) systems. These initiatives, going back over 55 years, resulted from federal budgets that inadequately link programs to costs.[2, 3] Because this problem has not yet been "solved"—nor is it easily solvable—we have experienced the institution of numerous PBM systems. Each PBM system attempts to provide an answer to a nearly 70-year-old question asked by Vladimir Orlando Key, an American political scientist: "On what basis shall it be decided to allocate X dollars to activity A instead of activity B?"

PBM systems are designed to focus the attention of an organization's personnel, and its management in particular, on desired outcomes—those results that can be directly mapped to strategic goals. Secondarily, PBM systems help managers to use resources to achieve those outcomes in a cost-effec-

DR. DIANA ANGELIS

DR. NATALIE WEBB

tive manner. These systems tend to connect performance measures through some sort of hierarchy, starting at the top with strategies, flowing down to useful metrics, and from relevant processes and their metrics, back to strategies.

As one might surmise from the number of PBM systems tried in government, public sector managers have found connecting resources to outcomes to be complicated, demanding, and data-intensive. Particularly challenging is the effort to find outcomes in the provision of public goods and services. The provision of national defense, education, public health and safety, and similar public services generally results in outcomes that are difficult to devise measures with which all can agree.

This article describes what can be measured; keys to developing good measures; which measures matter for determining effectiveness and efficiency; and the measures and data necessary to begin measuring what matters. While the article describes performance measurement in general, the framework and many examples focus on the Department of Defense (DoD) community.

A Framework for Understanding Defense Performance

Figure 1 presents a circular-flow model of DoD performance, emphasizing an ongoing and iterative process. We use this process to describe desirable measures—those that indicate how well an organization achieves its stated goals and desired outcomes (that is, whether it is *effective*) and whether it conducts its activities (that is, uses inputs to produce outputs) by employing the least cost combination of resources necessary (that is, whether it is *efficient*).

Beginning at the top with threats, the United States federal government develops defense policies and strategies based on an assessment of the threat environment. DoD leaders plan for forces needed to implement the policies and strategies and prepare budgets to resource personnel, equipment, and other inputs needed to provide defense activities.

Once appropriations are enacted, different types of funds flow to organizations through apportionments, allocations, and allotments. Each organization buys *inputs* (resources used, such as labor, materials, infrastructure, etc.) needed to undertake its own activities and missions. Activities combine the inputs to produce *outputs*.

> "*Efficiency* is the way we measure whether we are doing things right: Did we use the least cost combination of inputs in producing services or goods? *Effectiveness* answers the question, Are we doing the right things?"

Efficiency is the way we measure whether we are doing things right: Did we use the least cost combination of inputs in producing services or goods? This is reflected in the "inputs to activities to outputs" part of the model. Rather than focusing on line items and appropriations (budgeting processes), efficiency requires understanding the outputs, including the ability to measure outputs and assign them to the inputs used to produce them.

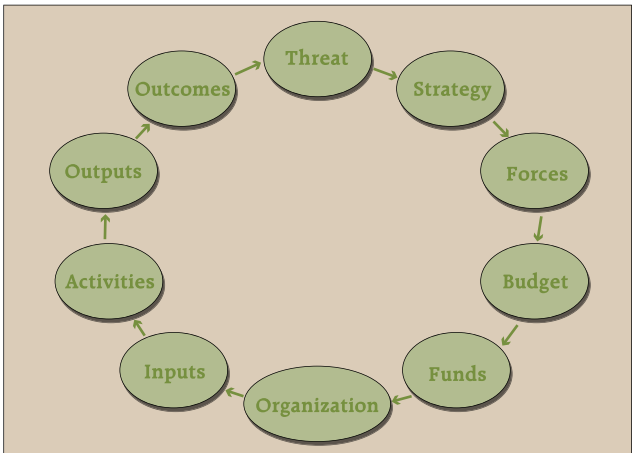


Figure 1. Circular Flow of Defense Performance

Effectiveness comes from the "outputs to outcomes" part of the model. It describes how well the organization's outputs performed relative to the strategies or high-level goals that the organization set out to achieve. Effectiveness also answers the question, Are we doing the right things?

Outcomes are results, consequences, effects, or impacts of direct importance to stakeholders. Outcomes, however, depend not only on the output generated but also on the interaction of the output with the environment and the interpretation of stakeholders as to the success of

"Often, particularly in the public sector's provision of goods and services, it is difficult to find a suitable natural measure. In such cases, we might use a proxy measure, that is, one that is related to the performance target but does not measure it directly."

that interaction. For example, attaining the goal of providing national security could be measured by how members of the public feel about whether they are "safe," whether "we are winning the battle," or "how many attacks occur daily." Clearly, there are many measures, each with some utility for a particular audience and none fully measuring the "success" of the goal of providing national security.

In the following sections, we provide definitions and attributes of good performance measures and then provide more detail on measuring efficiency and effectiveness.

Developing Desirable Measures

Desirable metrics flow from the goals and show whether a program or an organization is effective and efficient. Developing useful measures also requires some understanding of types and characteristics of good measures. When he described the measurement of objectives in his book *Value-Focused Thinking*, Ralph L. Keeney identified three types of measures: natural, constructed, and proxy. A natural measure is one that is in general use and is easily understood because it has a common interpretation. For example, the U.S. dollar is a natural measure for cost. Natural measures directly gauge the degree to which a performance target is met. For example, if the objective is to minimize cost, then we can directly measure how well we achieve that objective in dollars.

Often, particularly in the public sector's provision of goods and services, it is difficult to find a suitable natural measure. In such cases, we might use a proxy measure, that is, one that is related to the performance target but does not measure it directly. For example, if we want to measure workplace safety, we might use the number of accidents as a proxy measure.

In many cases, we have no clear understanding of how performance should be measured, so we might use a constructed measure, one that describes different levels of achievement and assigns a numerical value to each level. To assess, for instance, a safety hazard, a simple constructed scale might be:

5 Fatal injury
4 Permanent disabling injury
3 Disabling injury
2 Injury causing time off work
1 First aid only

Note that constructed scales involve some degree of subjectivity; therefore, each level must be carefully described to reduce ambiguity.

Finding the right measures requires careful consideration of both the intended and the unintended consequences, as well as actual and perceived incentives: "Tell me how you are going to measure me, and I'll tell you how I will behave," as stated by world business leader Eliyahu M. Goldratt. In a 2005 article for *Operations Research*, Dr. Keeney and R.S. Gregory suggest five desirable properties to keep in mind as we develop measures. The following box identifies the properties.

Five Desirable Properties for Performance Measures

Unambiguous—The relationship between the measure and the performance target should be clear and easy to interpret.

Comprehensive—All possible levels of performance are covered by the measure.

Direct—The measure should directly reflect the desired performance.

Operational—The data being used to measure performance is available or can be obtained with a reasonable amount of cost and effort.

Understandable—Everyone agrees on what is being measured and how it will be measured.

To recap, the development and selection of performance measures are complex and difficult tasks. Measures of outcomes try to link outputs (the product of activity) to outcomes (the results desired). Measures of efficiency link the cost of inputs to outputs. Measur-

ing performance in the public sector generally requires the use of proxies to substitute for direct measures of outcomes and/or constructed measures related to desired outcomes and may require similar measures of outputs. In all cases, those setting measures should try to make their measures meet the five desired characteristics to the extent possible. Finally, good measures should provide decision makers with useful information while encouraging desirable behavior that contributes to the achievement of goals and objectives.

Measuring Efficiency

As previously defined, efficiency is a measure of how well an organization turns inputs into outputs. To measure efficiency, we must understand the relationship between the *cost* of inputs and the *amount* of outputs. Unit cost or average cost is a commonly accepted measure of efficiency that is useful for linking performance to budgets. It is defined as cost divided by units. Calculating the unit cost of an output requires information about the following:

- The total cost of the inputs used to produce the outputs
- How the inputs are used to produce the outputs
- The amount and type of outputs produced

We can easily find the cost of inputs procured by an organization by looking at budget and accounting records. To find the total cost of inputs used by the organization to produce an output, however, we must look at how those outputs are produced, that is, the processes and activities of the organization.

In most cases, organizations use some inputs that are not funded from their operating budgets. For example, military labor is an essential input to the activities of the organization and the production of outputs; however, we fund military personnel entitlements from Active and Reserve component appropriations that are generally not controlled at the organization level.

To determine how inputs produce outputs and the costs of outputs, we use cost accounting, which is the process of collecting and allocating costs to outputs. Cost-accounting systems accumulate costs by components of the organization responsible for producing goods and services and allocate them to the outputs produced within each component.

Simply dividing the total expenditures of an organization by the number of outputs produced provides a useless number unless we divide outputs into relatively homogeneous groups. For example, the unit cost of educating traditional students differs significantly from the unit cost of educating special-needs students. In general, outputs with similar characteristics require similar activities. Thus expenditures on inputs must be accumulated separately for each type of output, and outputs must be divided into relatively homogeneous groups so that the unit cost within each group is comparable and relevant to the management of the organization's efficiency.

One of the most difficult aspects of calculating unit cost within service organizations involves the identification of exactly what constitutes a unit of output. Compounding this problem is the fact that input costs are usually measured for a period of time (normally a year); outputs may take a longer period to be produced. If we intend to manage efficiency with unit cost, we must address these problems.

In summary, unit cost is an easily understood measure of efficiency, but calculating a reliable value requires some analysis. Standard budgeting and accounting systems are designed to measure input costs and do not directly provide the necessary information. Cost-accounting systems can be designed to provide useful information but can be expensive and time-consuming to implement. For managers, the key is to develop a measure of efficiency that is both reliable and useful.

Effectiveness Measures

Effectiveness, as we stated, describes how well the organization's output achieves the strategies or high-level goals of the organization. To measure effectiveness requires knowing

- Who are the customers and/or stakeholders?
- What are their expectations?
- What are their desired outcomes?

Proper construction of effectiveness measures requires connecting metrics to organizational goals. Because outcomes are subjective and

"To measure efficiency, we must understand the relationship between the *cost* of inputs and the *amount* of outputs. One of the most difficult aspects of calculating unit cost within service organizations involves the identification of exactly what constitutes a unit of output."

"Finally, it is important to question what we measure. Where measures contribute to achieving results, we should use them; where they do not, measures should be abandoned."

the future is unknown, measures of effectiveness can be difficult to define and use with accuracy. In the DoD community, we often substitute input or output measures for outcomes. To the extent that these measures cause a desired result, they may be reasonable proxies for the desired outcome. Note, however, that at least part of an outcome results from the interaction of the output with an environment.

No matter how efficient an organization is in providing the output, other factors affect the outcome. For example, the output "x" sorties flown may result in winning a battle or losing an aircraft; both outcomes depend on many factors including those beyond the control of the squadron, pilot, or commander. Even the outcome of "winning a battle" may not be universally agreed upon by stakeholders. Similarly, the outcome of losing an aircraft depends on where you sit—were you the pilot, the squadron commander, the combatant commander, the taxpayer, or another interested party?

Even in noncombat organizations, effectiveness can be difficult to assess. Meals served (an output) may be tasty, delicious, fattening, and unhealthy. Each outcome is acceptable to some people (perhaps even desired by some!) and not to others. Well-prepared customer service surveys can provide data that assess (subjectively) whether desired outcomes were met. In the preceding example, as in the provision of many public services, conflicting goals can affect survey results—some folks may want healthy meals while others want something else. Customers will have different responses to the survey based on their desires, their knowledge of the relative healthiness of their food choices, etc.

Perhaps one of the most interesting and difficult output/outcome dilemmas in the DoD community comes from measuring readiness. Readiness may be the end state of a particular organization's efforts. If completing a standard level of maintenance (such as training to a standard, equipping, and assigning personnel together provide the end result of providing assets ready to fight), an organization may have achieved its desired outcome. A combatant commander, on the other hand, requires the asset as an input to the process of going into battle. Thus one person's outputs and outcomes may be only intermediate outcomes or even inputs to the organization as a whole.[4]

This is true in other aspects of organizations as well. The output of cutting checks may result in personnel receiving accurate paychecks on time—a positive and expected outcome for those receiving the paychecks, but only an intermediate outcome (at best) for DoD leadership, whose immediate concern most likely is the results of combat and other operations-oriented actions. (Of course, paying personnel on time has repercussions that can affect higher-level strategic goals, so such an issue may rise to a first-order performance failure in the minds of DoD leadership.)

In summary, informative measures of effectiveness link cause and effect. In the many cases where effectiveness cannot be directly measured in the DoD community, we use proxy (often output) measures (such as sorties flown, bombs dropped, or casualties). To the degree that these measures predict success (and none of them is likely to do so completely without some other measure of political, environmental, or other factors), they may be meaningful. When a composite measure, such as readiness, appears to offer better predictive power, it may be substituted for a nonmeasurable outcome.

It is also more difficult (and generally almost impossible) to assign the costs of inputs to outcomes. Where a relationship can be shown (such as lower funding resulting in overall readiness over time), perhaps using data for a funding request may be appropriate. In general, though, such data have tenuous linkages at best and may provoke bad behavior. (An example of bad behavior results from a belief that more money will fix everything—therefore, spend it all and ask for more rather than prioritizing repairs, asking hard questions about what we do not need to do to achieve the goals, etc.)

Finally, it is important to question what we measure. Where measures contribute to achieving results, we should use them; where they do not, measures should be abandoned. And using measures for a purpose not intended provides extremely bad outcomes for the organization. For example, using obligation rates to evaluate executing a budget may make sense in some contexts but likely makes little sense in evaluating whether the program executing the budget is conforming to its planned procurement.

One last word of caution: All of us should be careful to explain fully our use of performance measurement terms. "Effects-based" simply means

considering the results or outcomes. The same is true of any number of current buzzwords. (One of the authors' least favorite is the current use of ROI to discuss return on expenditures for operations—this makes no sense. Returns on investment means just that—returns to investment spending!) Thinking carefully about what you need to measure to achieve a goal (effectiveness measures) and how to use resources to achieve them in the most cost-effective manner (efficiency measures) can help avoid a multitude of problems in measuring performance.

Summary

We have presented a framework for DoD performance measurement. The framework shows how performance measures are linked to strategic goals and objectives. The process begins with threat analysis leading to the development of policies and strategies. The DoD identifies the capabilities (forces) needed to support the policies and strategies that guide the development of the budget necessary to acquire those capabilities. The budget leads to appropriations that provide funds to organizations responsible for providing DoD capabilities. Funds are used to buy inputs, and inputs are combined in activities to produce outputs. This part of the framework forms the basis for the measurement of efficiency—how well we turn inputs into outputs. Effectiveness comes from the output-to-outcome part of the model: Did the outputs produce the desired effects? If we achieve the desired effects, then the threat is reduced and the cycle is repeated.

To measure efficiency, we must understand the relationship between the *cost* of inputs and the *amount* of outputs—and cost accounting can help us do this. To measure effectiveness, we must understand the relationship between the organization's goals and objectives and its outputs—the outcomes. Because outcomes are subjective and the future is unknown, measures of effectiveness can be difficult to define and use with accuracy. Finding the right measures requires careful consideration of both the intended and the unintended consequences as well as actual and perceived incentives. By using the framework presented in this article, you can develop DoD performance measures that are linked to goals and objectives. ■

DR. NATALIE WEBB

Dr. Natalie Webb is an associate professor of economics at the Defense Resources Management Institute (DRMI), Naval Postgraduate School, Monterey, California, where she has been on the faculty for more than 16 years.

DR. DIANE ANGELIS

Dr. Diana Angelis is an associate professor of industrial and system engineering at DRMI and has been at the Institute for 12 years.

ENDNOTES

[1]These also include Criteria-Based Assessments, Managing for Results, Performance Contracting, Total Quality Management, and Lean Six Sigma; however, some of these, particularly the latter, were designed to deal with production efficiencies rather than program results.

[2]See, for example, the first and second Hoover Commissions of 1947 and 1953, respectively. Available in the U.S. Commission on Organization of the Executive Branch of the Government, Budgeting and Accounting (Washington, D.C., 1949), p. 8.; see also U.S. Commission on Organization of the Executive Branch of the Government, Budgeting and Accounting (Washington, D.C., 1949), p. 8.

[3]U.S. General Accounting Office, Performance Budgeting: Past Initiatives Offer Insights for GPRA Implementation, GAO/AIMD-97-46 (Washington, D.C., March 1997), p. 33.

[4]Outcomes definitions adapted from Harry P. Hatry (2002) "What performance measures should be tracked?" in Dall W. Forsythe, ed *Quicker, Better, Cheaper: Managing Performance in American Government* (New York: Rockefeller Institute), p. 18.

WHAT ARE WE TALKING ABOUT?

Sometimes we do things differently in the financial management community. COMPtroller, CONtroller, COMtroller—we have all heard it pronounced these ways. Well, which one is right? Merriam-Webster's dictionary and Bartleby.com say all three. And, yes, I know that last you checked Merriam-Webster and Bartleby are not in your chain of command. But just in case you are curious, this is what the sources say:

Merriam-Webster says:
comp·trol·ler, Pronunciation: \kən-ˈtrō-lər, ˈkäm(p)-ˌ, käm(p)-ˈ\ Date: 15th century

1: a royal-household official who examines and supervises expenditures 2: a public official who audits government accounts and sometimes certifies expenditures 3: Controller

Bartleby.com says:
"This word is first recorded in the 15th century as an alternate spelling for *controller*, the first syllable of which had become associated with the etymologically unrelated word *count* and its variant *compt*. Although the historical pronunciation for this word would be the same as for controller, evidence indicates that the spelling pronunciations (kŏmp-trō´lər) and (kŏmp´trō´´lər) are probably now used by a majority of speakers. In a recent ballot, 43 percent of the Usage Panel indicated that they pronounce *comptroller* like *controller*, while 57 percent pronounce it as it is spelled, with stress on either the first or second syllable. And of those who say they pronounce *comptroller* like *controller*, about half indicated that they also consider one or the other of the spelling pronunciations acceptable."

So, just appreciate this situation as one of those times when everyone is right and pronounce comptroller however you prefer. ■