SELECTING EFFECTIVE ACQUISITION PROCESS METRICS

Aron Pinker, Charles G. Smith, and Jack W. Booher

Metrics for assessing the acquisition reform process are now being actively sought by the DoD. It is difficult to identify meaningful metrics that can be conveniently calculated. Using our experience from the Partnership Process for Electronic Warfare (EW) Acquisition, we describe a reasonable approach to effective selection of metrics. We examine DoD initiatives aimed at measuring acquisition reform, identify a process for establishing metrics, suggest a basis for ordering metrics, and provide examples of metrics.

This article is a result of the Secretary of the Air Force Electronic Combat Division's (SAF/AQPE's) effort to design a new approach to the acquisition of Electronic Warfare (EW) systems. SAF/AQPE assembled an EW Acquisition Partnership team to design an acquisition process that seamlessly integrates the warfighter's requirements with product development and testing. From its inception, the EW team recognized that improving the EW systems acquisition process requires identification of the baseline acquisition process for EW systems and definition, or development, of a new acquisition process. To attain this objective and demonstrate that improvement has been achieved, it is imperative to have some measures, or metrics, for comparing the old process (baseline) with the new process (acquisition reform). Here we present some of our insights on metrics that could be useful to the DoD acquisition community.

PROBLEMS IN DEVELOPING METRICS

Most people who work with metrics recognize that it is not easy to identify meaningful metrics that can be conveniently calculated (Dellinger, 1994).

"Metrics allow us to baseline where we are, identify the impediments to the process, and track the impact of management actions on processes and other process changes."
–Gen. Thomas R. Ferguson, Jr.

The main consideration in Air Force acquisition reform is whether the new process enables us to field *better* weapon systems, *faster*, and *cheaper*.

The first problem with developing metrics for the acquisition process is that we cannot directly measure these attributes. So they are useless as metrics; we must use other, quantifiable, "surrogate" metrics instead. But it is not easy to decide what these surrogate metrics should be, and it is not always clear how they would contribute to the goal of fielding military systems that are better, faster, and cheaper.

A second problem with formulating metrics is the fact that a weapon system acquisition takes place over a long period of time. The success or failure of the acquisition is determined in retrospect by how well the weapon system has served the military. Consequently, we can assess the success of the acquisition process only in a post mortem. Such an assessment would, of course, be of merely historical interest and little practical use. We will later suggest a method for creating a top-down (or bottom-up) hierarchy of metrics that links surrogate metrics to the true metrics by means of Quality Function Deployment (QFD) (Fortuna, 1988) and the Analytic Hierarchy Process (AHP) (Saaty, 1980).

DEVELOPING STANDARDIZED METRICS

As the defense acquisition system is being streamlined, DoD is also considering ways to measure the improvement as it occurs. Measuring improvement starts with identifying the changes in process brought about by acquisition reform and providing a comprehensive plan for estimating and measuring these changes. Dr. Paul Kaminski, Under Secretary of Defense (Acquisition and Technology) believes that the Pentagon should have Defense Department-wide metrics (Meadows, 1995). If this standardization is achieved, it would provide a useful basis for comparing the various acquisition reform initiatives.

This raises the question of which metrics can be shared by all commands and which would only apply to specialized activities. For example, while the EW acquisition community may have some metrics that are shared by the general DoD procurement community, EW may have some unique service-specific or area-specific metrics. Additionally, if different

Aron Pinker holds a M.Sc. in physics, mathematics, and meterology from Hebrew University, and a Ph.D. in mathematics from Columbia University. For a number of years he was a professor of mathematics at Frostburg State University. Dr. Pinker is the author of several books on physics, and numerous papers on mathematics, physics, and decision sciences. He produces the weekly newsletter *GPS* for *Air Power*.

Charles G. Smith holds a B.S. in industrial management from Syracuse University and a B.S. and M.S. in electrical engineering from the Air Force Institute of Technology. For more than three decades he has been involved in the DoD acquisition of avionics systems and support equipment. Mr. Smith has authored numerous papers on avionics standardization, logistics, and technology.

Colonel Jack W. Booher, USAF (Ret), is the former division Chief of the Electronic Combat Division at the Office of the Assistant Secretary of the Air Force (Acquisition), Directorate of Global Power Programs. Mr. Booher is currently employed by Lockheed-Sanders.

Table 1. Initial Metrics

TYPE OF METRIC	METRIC
Cost	Consumable item price index, military specification conversion price benefit
Acquisition performance	Contract defaults, contract changes
Schedule	Acquisition phase time, administrative lead time, multiyear procurements; FACNET transactions, logistics response time
Commercial practices	Contract specifications, credit card purchases

commands have different senses of mission criticality, they would weight the shared metrics differently.

Consequently, using our experience from the EW Acquisition Partnership, we intend to describe a reasonable approach for selecting metrics. In the following sections we will examine DoD initiatives aimed at measuring acquisition reform, identify a process for establishing metrics, provide examples of metrics, and suggest a basis for selecting and ordering metrics.

Recent DoD Initiatives

Process Action Teams (PATs)

Last year Dr. Paul Kaminski and Colleen A. Preston, former Deputy Under Secretary of Defense for Acquisition Reform, chartered several process action teams (PATs) to recommend actions for reforming DoD acquisition practices and to define metrics for assessing the effectiveness of the recommended reforms. The PATs were fairly successful in identifying simplifications and improvements to reform DoD acquisition practices. The definition of metrics, however, has turned out to be a major difficulty. The PATs have struggled to come up with at least some metrics. Yet, they never explained the interrelationship and connection of these metrics to the over-all goal.

The Tiger Team

After the PATs' attempts at defining metrics, the Defense Standards Improvement Council formed a metrics Tiger Team, led by the Office of the Assistant Secretary of the Army, Research, Development, and Acquisition (OASA/RDA), Acquisition Reform Office, to develop metrics and a method for collecting data for these metrics. This team has proposed a set of initial strategic outcome metrics for measuring the impact of acquisition reform. Preston has approved the strategic outcome metrics in Table 1 and has authorized the OASA/RDA to collect the necessary data.

It appears that the Tiger Team selected these metrics because they are relatively easy to collect. From the warfighters' perspective, the category of "system performance" has been omitted. Also, the Tiger Team has not addressed the issue of quick integration of advanced technologies. The categories of metrics in Table 1 will probably be expanded in the future.

"Because surrogate metrics are not true metrics, we need to know how strongly they represent the true metrics."

Appendix A presents a list of the strategic outcome metrics that have been suggested to the Acquisition Reform Senior Steering Group. Appendix B presents the algorithms for computing the initial set of selected metrics.

Aquisition Reform Benchmarking Group

On Sept. 18, 1995, Preston established the Acquisition Reform Benchmarking Group (ARBG), chaired by William E. Mounts from her office. The ARBG will receive, assemble, and assess data from these and other acquisition reform strategic outcome metrics. The group will also assess the suitability of other metrics proposed by the various acquisition reform PATs. Interim results can be found on the World Wide Web at:

http://www.acq.osd.mil/ar

Conceptual Considerations in Selecting Metrics

Defining True and Surrogate Metrics

Because our interest is in the acquisition process, we have chosen to define metrics for this process rather than metrics in general as shown in Table 1.

Acquisition reform metrics are the numerical values by which we gauge progress toward meeting acquisition reform objectives.

If the overall objective of the acquisition reform is to field *faster*, *better*, and *cheaper* weapon systems, then a true metric would be any numerical value that enables us to assess *how much* faster, *how much* better, and *how much* cheaper a given acquisition process is. Unfortunately, we do not have such true metrics; we do not know how to directly measure these qualities. The terms *faster*, *better*, and *cheaper* have so many possible meanings that we must restrict these terms to some of their more specific characteristics. To do this we have to use "surrogate metrics."

A *surrogate metric* is a measurable characteristic of the acquisition process that presumably reflects the behavior of a true metric.

Because surrogate metrics are not true metrics, we need to know how strongly they represent the true metrics. Moreover, some metrics may be better described as submetrics that together constitute a higher level. This grouping leads to a hierarchical structure of metrics with the many surrogate metrics at the bottom and a few true metrics at the top. This grouping also requires us to determine how the lower-level metrics contribute to the higher-level metrics.

Brainstorming Potential Metrics

One can usually gather many potential metrics for a process. Follow these guidelines to brainstorm for potential metrics:

1. Identify the specific segment of the process that is to be evaluated.

2. Identify the pertinent properties of what is to be measured.

3. Identify types of potential metrics.

4. Select a few metrics and provide a rationale for the specific selection.

5. Find bounds on what is being measured.

AVOIDING INEFFECTIVE METRICS

Once you have discovered several potential metrics, determine which ones will be most useful. A good metric will be meaningful, logical, simple to express, understandable, repeatedly and quickly derivable, unambiguously defined, and derivable from economically collectible data.

In addition, a good metric will indicate trends, suggest corrective actions, and numerically describe the progress toward the objective.

While it is important to be able to identify a good metric, it is also important to know what is *not* a metric. Metrics are not charts, schedules, goals, objectives, strategies, plans, missions, guiding principles, counts of activity, single-point statistics, or rankings. Also, tracking a process is not necessarily the same as tracking a metric. In spite of this, one IPT suggested using the following measurements as metrics:

> "Program managers should track use of military unique specifications and standards and report out at milestone/program reviews" (OUST[A&T], 1994, p. 53)

> "The Standards Improvement Executives shall be responsible for tracking implementation of all acquisition reform issues related to specifications and standards" (OUSD [A&T], 1994, p. 165).

Another IPT suggested that contractor responses to a questionnaire would serve as an input to a database, which would eventually be used for developing metrics. This proposed questionnaire included the following questions (OUSD[A&T], 1994, p. 27):

1. Are there any military specifications or standards required as a part of this solicitation which could be better served by a commercial specification?

2. Were any changes required in your routine manufacturing process specifically to accommodate this DoD purchase? Do you believe that the changes added value to the product?

3. Did you offer alternatives to requirements of any military specifications or standards? Do you feel that your alternatives were given adequate consideration by the procuring agency? Were any adopted?

4. How would you improve the solicitation to allow you, and other contractors, to quote a lower product cost while maintaining identical product performance?

Such questionnaires may solicit ideas for reform, but they seem to have little value for forming metrics, because they do not call for numerical, quantifiable responses. Could questionnaires be used for developing metrics? Could they be used to provide metrics that are of immediate use? The answers to these questions are not easy and may depend on the particular program for which metrics are developed.

THE REQUIRED DETAILS FOR A METRIC

The definition of a metric tends to be simple, because a metric should be easy to explain and calculate. Yet, from the technical point of view, many details of each metric must be specified to ensure commonality of the derived metrics. For each metric, at least, the following details must be specified:

1. Description of the population that the metric includes.

2. Identification of the source of data.

3. Precise definition of key terms.

4. Statement of the mathematical expressions that will be used to derive various values.

5. Specification of frequency of measurements to derive the metric.

6. Description of the graphics that will be used to display the data.

7. Specification of user's tolerance levels (i.e., "control limits").

8. Listing of desired outcomes expressed in terms of a positive or negative trend (not a numerical goal).

9. Linkage between the metric and the activity being measured.

10. Linkage between the surrogate metric and the true metric.

CREATING A METRIC

Having laid out general guidelines and requirements for designing metrics, we now describe a step-by-step procedure for establishing and using metrics to assess improvement in the acquisition process. Follow this procedure to create a metric:

Identify the purpose of the metric. The purpose of the metric should reflect the purpose of the acquisition reform initiative and its mission, vision, goals, and objectives.

Develop an operational definition of the metric. Define the who, what, when, why, and how of this metric in sufficient detail to permit consistent, repeatable, and valid measurement of the acquisition process.

Examine existing means of measuring. Check whether existing metrics or process measuring means could be adapted to satisfy the operational definition of the metric. In other words, do not "reinvent the wheel."

Generate new metrics. In the past, most metrics were not process-oriented; they were usually related to final outputs, products, or services. The focus is now on improving the new acquisition process so that superior final outputs are obtained. Currently, the underlying assumption for generating metrics is that by monitoring changes in the process we can assess process improvements.

Conduct a "goodness of fit" check. Check whether the newly generated metric satisfies the previously stated attributes of a good metric. Make sure that all the previously stated details can be provided for this metric. Check objectivity of the metric to ensure that the measurements or observations do not affect the outcome.

Choose a mode of display. Decide on the mode for presenting the metric. This decision will affect data collection and availability.

Conduct a "sanity" check. Acquire data for deriving the metric. Derive the metric for various instances and ask the customer to judge whether the metrics are meaningful. Does the metric measure what it is supposed to measure? Do the metric values correspond to intuition? If the answer is uncertain, return to the second step.

Form a consensus. Obtain consensus or buy-in from participants.

Create a database. Collect and analyze the metric's data over time and for different cases. Examine trends. Can you adequately explain counterintuitive metric values? For what lengths of time does the metric stabilize (i.e., does not deviate significantly from its mean)?

Communicate the metric. Be open to constructive criticism. Be ready to make adjustments.

Employ the metric. Metrics are just a means to an end—continuous process improvement. If there is confidence in the metric, then it should be used; otherwise, look for a new metric. Employing the metric allows you to refine it and make it an even better tool. (AFSC, 1991).

EXAMPLES OF METRICS

The following illustrate metrics at various levels of abstraction and areas of interest to the acquisition process. These metrics were collected from various sources, but most of them fall into the following categories: cost, acquisition performance, schedule, commercial practices, weapon system performance, and technology innovation.

Program office overhead. Program overhead as a fraction of total program cost.

Specifications conversion. Number of military specifications that have been replaced with industry standards.

Specifications elimination: Number of military specifications that have been eliminated; or, reduction in number of specifications and standards specified in a contract.

Cost and pricing data. Percentage of competitive, negotiated procurements requiring certified cost and pricing data; or, ratio of the number of contract awards with cost and pricing data to the total number of contract awards.

Funding stability. The number of times a program changes in terms of quantity or cost, due to fiscal pressures external to program executive officers (or an equivalent management level).

Program cost. Change in program cost as a consequence of changed acquisition processes.

Unit production price. Change in unit production cost as a consequence of changed acquisition processes.

Unit life-cycle cost. Change in projected unit life-cycle cost as a consequence of changed acquisition processes.

Operational performance versus cost. Compare operational test results versus specified performance for accuracy and reliability with Average Unit Production Price Milestone I cost analysis improvement group estimates versus contractor production proposals.

Commercial practices. Compare business-as-usual versus commercial practices costs.

Billing. Effect of milestone billing versus cost billing.

Oversight. Number of oversight personnel per program budget size.

Cost of performance. The kind of system performance that can be bought for a given cost. To derive this metric it would be necessary in some way to quantify various combinations of system performance. This is a formidable task open to controversy.

Commercial componentry. Percent of commercially available componentry: dollars of commercial material to dollars of total obligation.

System gestation time. Time for a system or item to progress from concept exploration and definition to start of production and deployment phase.

Contractor's past performance. Contractor's ranking relative to other contractors on a predetermined set of criteria.

Government-unique terms. Proportion of government-unique terms and conditions to total number of such terms and conditions in a contract.

Protests. Number of bid protests per number of bidders.

Regulatory cost premium. DoD cost premium (%) equals contractor compliance costs ($) divided by value-added costs ($) x 100.

Value-added costs. Value added costs as percent of total costs where value-added costs equal total costs minus costs of material purchases, including subcontracts minus profit minus corporate general and administrative allocations.

Contractor overhead. Compare percentage of direct and indirect costs for top defense contractors as a group and individually over time. Use the ratio of percent indirect costs to percent direct costs or dollars of indirect costs to dollars of direct costs.

Consumable item price index. Cost of a Defense Logistics Agency (DLA) predetermined set of consumables.

Contract defaults. Number of contract action defaults divided by the total number of contract actions.

Contract changes. Number of contract changes divided by the total number of contracts.

Contract protests. Number of protests resolved using the alternative dispute resolution process, and the number of protests that go to GAO and the General Service Board of Contract Appeals (GSBCA).

Administrative lead time. The average time from the signed formal requirements document to contract award.

Production lead time. Time from contract award to acceptance of first item or delivery.

Engineering changes. Number of engineering change proposals by program phase (demonstration and validation, engineering and manufacturing development, production startup).

Alternative specifications and standards. Number of contractors offering alternatives to military specifications and standards per 100 proposals.

Alternative specifications and standards with incentives. Percentage of solicitations resulting in incentive contracts where alternatives to military specifications and standards are offered.

Dissemination time. Time for processing and dissemination of requests for proposal, statement of work, and specifications and standards.

Degree of use of simulation and modeling. Percentage of contracts over $5 million using simulation and modeling to achieve cost performance tradeoffs.

Degree of activity-based costing and management. Percentage of contracts and contractors that use activity-based costing and management. (Activity-based costing identifies each category of cost [direct or overhead] and relates it to the specific product [e.g., military specification or standard, statement of work task, etc.] or product line that causes the activity to be needed and performed.)

Marginal ownership cost. Cost divided by operating time.

Technology gestation. Time from technological innovation to operational system integration.

Cost as an independent variable. Savings in a program when cost is used as independent variable.

Operational goals. Probability of achieving or exceeding stated operational profiles in a specified regime.

Reliability goals. Probability of system's satisfactory operation in given conditions.

Maintenance goals. Proportion of maintenance activities requiring a given level of maintenance.

Integrability. Ease of integrating the new system into an existing frame or organizational unit.

Mean time between failures (MTBF). The average number of operating hours between system failures.

This list is not exhaustive. Though most of these metrics have been taken from DoD sources, it is not clear whether all of them could serve as metrics for a specific service's acquisition process. Still, these examples provide some insights into the types of metrics that are being considered and the levels of abstraction that are needed.

Quality Function Deployment and Analytic Hierarchy Process

The metrics in the preceding section are clearly at different levels of abstraction. For instance, "commercial practices" is more abstract than "oversight," and "integrability" is more abstract than "mean time between failures." Metrics that are not very abstract are usually easier to measure. Yet, highly abstract metrics are often more useful for assessing a process. We propose using the analytic hierarchy process (AHP) and quality function deployment (QFD) to create a top-down (or bottom-up) hierarchy of metrics that translates *what we can measure into what we are interested in measuring.*

Using QFD and AHP to Weight Metrics

QFD is a structured process that facilitates a team approach to identifying and prioritizing customer requirements and translating these requirements into appropriate company requirements at each stage of the product life cycle—from research and development to manufacturing and support.

QFD's structured process consists of a set of interrelated matrices. These matrices are constructed by starting with the general goals (the "whats") that are to be achieved and then selecting the various means (the "hows") for achieving those goals. In the next step, the current "hows" become the goals (i.e., the "whats") and new "hows" are identified for achieving the new "whats". This process is repeated as many times as necessary to reach a desirable level of detail.

For any one matrix, the elements of the matrix are intuitive ratings (using the scale *weak = 1, medium = 3, high = 9*) of the contribution of a column means ("hows") to a row goal ("whats"). The relationship between the matrices is twofold: the columns (HOWs) of a matrix become the rows (WHATs) of the subsequent matrix, and the computed weights of a matrix's column become the weights of the subsequent matrix's rows. Thus, if a matrix entry is m_{ij} and the row weights are w_i then the computed column weights are:

$$\sum w_i\, m_{i1}/N,\ \sum w_i\, m_{i2}/N,\ \sum w_i\, m_{i3}/N,$$
$$\sum w_i\, m_{i4}/N,\ \ldots \sum w_i\, m_{ik}/N,$$

WHAT \ HOW	COST	ACQUISITION PERFORMANCE	SCHEDULE	COMMERCIAL PRACTICES	SYSTEM PERFORMANCE	TECHNOLOGY INNOVATION	
BETTER EW SYSTEMS	3	3	1	1	9	9	.40
CHEAPER EW SYSTEMS	9	3	3	3	1	1	.35
FASTER FIELDED EW SYSTEMS	9	3	3	3	1	3	.25
	.29	.13	.10	.10	.18	.21	

WHAT \ HOW	CONSUMABLE ITEM PRICE INDEX	MILITARY SPECS CONVER. ADVANT.	CONTR. DEFAULTS	CONTR. CHANGES	ACQUISITION PHASE TIME	ADMINISTRATIVE LEAD TIME	MULTI-YEAR PROCUREMENTS	FACNET TRANSACTIONS	LOGISTIC RESPONSE TIME	CONTR. SPECIFICATIONS	OVERHEAD	ACHIEVE OPER. GOALS	MTBF	INTEGRABILITY	
COST	9	9	3	9	9	3	9	3	3	9	9	3	3	3	.29
ACQUISITION PERFORMANCE	1	3	9	9	9	3	9	9	3	9	1	1	1	1	.13
SCHEDULE	1	1	3	9	9	3	9	3	1	1	1	1	1	1	.10
COMMERCIAL PRACTICES	1	9	1	1	1	1	1	3	1	1	1	1	1	1	.10
SYSTEM PERFORMANCE	1	3	1	3	1	1	1	1	1	3	1	9	9	3	.18
TECHNOLOGY INNOVATION	1	1	1	3	1	1	3	1	1	1	1	1	3	3	.21
	.06	.09	.05	.12	.10	.04	.11	.06	.04	.09	.06	.06	.07	.05	

Figure 1. Quality Function Deployment (QFD) for Acquisition Metrics

where k is the number of columns and N $= \sum\sum w_i m_{ik}$ is the normalization factor. These weights become the weights of the rows for the subsequent matrix. Thus, at any given step the analyst can readily ascertain the relative contribution of a means to the end goals. An example of the QFD process for metrics is shown in Figure 1.

For the first matrix, the row weights could be derived using the pairwise comparisons of the AHP. The analytic hierarchy process, introduced by Saaty in 1971, applies a structured process in which only two factors are compared at a time (i.e., pairwise comparisons) to a complex problem that is broken up into manageable super structures forming a hierarchy (Saaty, 1980). The preferences in a pairwise comparison are denoted numerically by a scale of one to nine. One denotes equal preference; nine denotes extreme preference. The mathematical algorithm of AHP converts these pairwise preferences into rankings of relative importance for each level. AHP provides a framework for the selection of a preferred alternative in a context of conflicting criteria. For the mathematical procedures and details of the process see Saaty (1980).

Our implementation of QFD for a hierarchy of metrics consists of the following steps:

Step one: Form an IPT. Organize an integrated product team (IPT) consisting of members that adequately represent the concerned community with respect to the system under consideration.

Step two: Construct a hierarchy of metrics. Charge the IPT with developing a hierarchy of metrics. For instance, the IPT could start with the first set of "whats" consisting of "better system," "cheaper system," and "faster fielded system." For the "hows", the IPT could choose the areas that Preston believes warrants the definition of metrics and some areas that are of importance to the warfighter. Thus, the IPT's initial set of "hows" could consist of cost, acquisition performance, schedule, commercial practices, system performance, and technology innovation.

It is possible that an IPT would come up with a different QFD matrix. However, we believe that it is best to have as much commonality as possible within the DoD community. In the next QFD matrix the "whats" are cost, acquisition performance, schedule, commercial practices, system performance, and technology innovation. See Figure 1 for an example of a hierarchy of metrics that the IPT could create.

Step three: Derive initial weights. Assign a numeric value to the metrics for "better system," "cheaper system," and "faster fielded system" that corresponds to their relative importance. In this case the IPT must decide the relative importance of only three metrics. This can be done directly without using AHP. However, each IPT member should perform this assessment individually and the IPT should average the values using the geometric mean.

Step four: Fill out the first QFD matrix. Using the scale *weak = 1, medium = 3, high = 9,* each IPT member fills out the first QFD matrix by answering the question "How strongly does the particular 'how' reflect the particular 'what'?" The IPT average matrix entry is the geometric mean of the corresponding individual entries. Note that we assume that any "how" reflects any "what" to some degree, however weak. Thus, $m_{ij} \geq 1$.

Step five: Compute the weights of the "hows." If a matrix entry is m_{ij} and the row weights are w_i, then the computed column weights are:

$$\sum w_i m_{i1}/N, \sum w_i m_{i2}/N,$$
$$\sum w_i m_{i3}/N, \sum w_i m_{i4}/N,$$
$$\ldots \sum w_i m_{ik}/N,$$

where k is the number of columns and $N = \sum\sum w_i m_{ik}$ is the normalization factor. These weights are the weights of the rows for the subsequent matrix.

Step six: Repeat this process. Repeat steps four and five as many times as needed to develop a QFD structure that links the true metrics to measurable surrogate metrics.

Benefits From Applying QFD

Application of QFD to a hierarchy of metrics can provide the decision maker with significant insights. In the following we list some of these insights and illustrate them by means of the example in Figure 1 (see the first two matrices). Note that the entries in the QFD matrices of the figure are notional and that the rankings were derived using the QFD algorithms.

Application of QFD to a hierarchy of metrics provides indications of the metrics on which to concentrate the data collection effort. The first matrix in Figure 1 ranks the "hows" that contribute to the true metrics that measure the attainment of better EW systems, cheaper EW systems, and faster fielded EW systems. Metrics that pertain to schedule and commercial practices appear to have been ranked lowest (.10). This could indicate to a decision maker that his efforts should, perhaps, not be focused on these metrics if his resources are limited. The decision maker may wish to defer any such decision in the first step of the QFD process and wait to see the rankings in the second step, where a finer substructure is assessed. In this case, they will find that such metrics as contractor defaults, administrative lead time, and logistic response time rank very low. Now the decision maker is at a level of abstraction and detail that permits him to reconsider the collection of data for the measurable surrogate metrics.

Periodic application of QFD to a hierarchy of metrics provides a means for assessing the sensitivity of lower-level rankings to changes in higher-level rankings. Suppose that the DoD's emphasis has changed. Cheaper EW systems become more important (e.g., .40) and the faster fielding EW systems become less important (e.g., .20), while better EW systems remained as important as before. How would this change effect the ranking of the various metrics? The decision maker may *feel* that this change in emphasis is considerable and would affect the rankings of the metrics. However, the recalculation of "hows" rankings would show that they did not change, and change in current practices is not warranted.

Suppose that there is another change in DoD's emphasis. Cheaper EW systems become more important (e.g., .45) and the faster fielding EW systems become more important (e.g.,.45), while better EW systems become less important (e.g., .1). In this case the importance of the cost metrics jumps to .39 while that of system performance drops to .08. Clearly, the decision makers would have to make some adjustments and shifts in their efforts if they are constrained by budgets.

Application of QFD to a hierarchy of metrics provides a means for monitoring the relative importance of the metrics as a function of time. As time passes, the "whats," the "hows," or both may change. If the QFD computations are repeated at fixed intervals of time we may observe a shift in the importance of various metrics. Such indications could provide the decision maker with the necessary time to prepare to shift from one set of metrics to another.

> **"If the QFD computations are repeated at fixed intervals of time we may observe a shift in the importance of various metrics."**

Application of QFD to a hierarchy of metrics provides a means for comparing the metrics for the acquisition of two distinct weapon systems. Suppose that the QFD matrices in Figure 1 were obtained for an airframe and a similar set of matrices (with different entries and initial weights) was obtained for the avionics of this airframe. The differences in the rankings could then provide the decision maker with interesting information on potential problems with integration of the two systems.

Application of QFD to a hierarchy of metrics provides a link between what we can measure and what we are interested in measuring. Figure 1 indicates that the metric "contract changes" (.12 weight) is more strongly linked to the true metrics better EW systems, cheaper EW systems, and faster fielded EW systems, than the metric "logistic response time" (.04) weight.

UNDERSTANDING THE USE OF METRICS

STATISTICAL SIGNIFICANCE OF CHANGE

Suppose that we have identified a metric for a task or a process. Next, suppose that this metric is higher for the current process than for the previous process, and higher is better. Does this mean that the current process is better than the previous process? Not necessarily.

First, we must prove that the difference is statistically significant. However, this requires us to have a sizable sample of similar cases, make assumptions about the population probability distribution, and choose the statistics that will be used. Unfortunately, this kind of information is not available for military-unique programs. Consequently, we can rarely say with certainty that a positive change in a metric indicates a real improvement. We could say that an improvement was achieved only if the change in the metric was spectacular.

We have already noted that many metrics can be usually defined for a task or process. If all these metrics are independent and point in the positive direction, then we would be more certain that a positive improvement had occurred. On the other hand, if the metrics point in different directions, we cannot form any definite conclusion.

Measuring Variations

While a set of metrics often does not allow us to draw solid conclusions, the comparative metric values could still benefit the acquisition process by indicating process variations that cause unsatisfactory performance. Thus, linkages between metrics of the left column and metrics of the right column in Figure 2 could serve as indicators of process variations. However, such linkages can be established with reasonable confidence only if data from repeatable processes is available.

Repeatable processes, highly desirable for drawing statistical inferences, are usually unavailable for a weapon system acquisition. Such an acquisition is in most cases a unique event. Moreover, the success or failure of the acquisition is determined in retrospect by how well the weapon system has served the military. Consequently, we can assess the success of the acquisition process only in a postmortem. Our challenge is to define metrics that show how the pieces of the process are doing and then to extrapolate these results to the entire process.

Conclusions

We have outlined requirements and procedures for defining meaningful metrics for the acquisition process. It does not provide a prescription for the generation of metrics, however. For some processes useful metrics come readily to mind, for others one must employ substantial insight and creativity. But in each case the methods presented should be of practical use.

Since directly computable metrics tend to be limited in scope and specific in nature, we need to know how to combine the various metrics into one big picture. We propose to accomplish this by application of QFD or AHP. In these processes we do not actually combine the metrics,

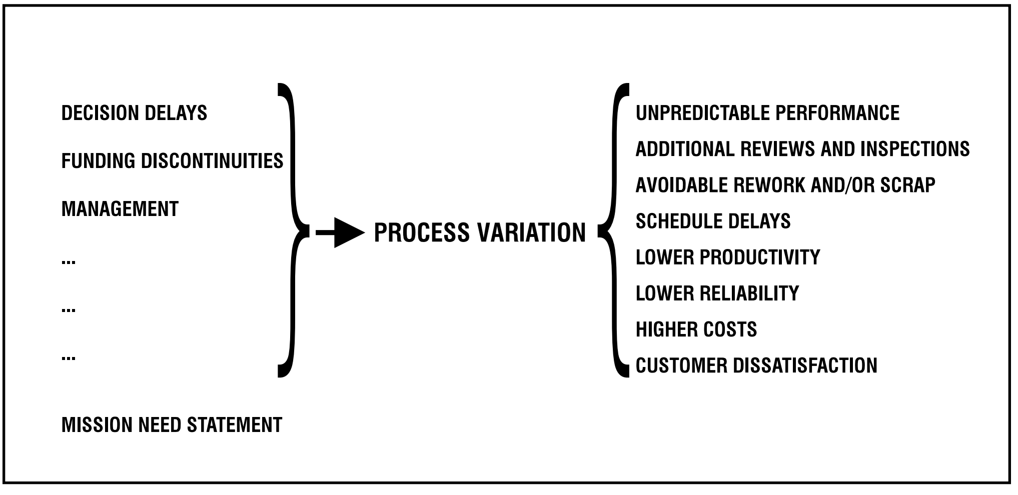

Figure 2. Causes of Process Variation

but rather gauge the relative contribution of a metric at one level to the metrics at a higher level. This seems to be a more prudent path to follow, because past efforts to combine metrics numerically have usually failed.

QFD offers interesting opportunities for linking metrics from one level of abstraction to a higher level. The ranking of metrics in the QFD process allows one to select the relatively important metrics. This prioritization could lead to more efficient strategies for assessing the acquisition process.

Definition of a metric is the beginning of a process of continual refinement for measuring process outcomes. As data for a metric is collected and the metric is used, much is learned that could shape an even better metric. Efforts to define useful metrics for the acquisition reform process must focus on measures that give insights into reform effects, not specific acquisition program indicators. Employing the methodology outlined in this paper should help to keep the focus where it ought to be—on reform processes.

ACKNOWLEDGMENT

We thank Maj. Art Huber for his constructive remarks.

REFERENCES

Air Force System Command. (1991, August). The Metrics Handbook (p. 3-3). Andrews Air Force Base, MD.

Dellinger, L. (1994, November–December). Secretary Preston addresses DAU training and education conference. Program Manager, 21.

Fortuna, R. M. (1988, June). Beyond quality: Taking SPC upstream. Quality Progress, 23.

Meadows, S. I. (1995, July/August). Army grapples with means to appraise reform efforts. National Defense, 31.

Office of the Under Secretary of Defense for Acquisition & Technology. (1994, April). Report of the Process Action Team on Military Specifications and Standards. Washington, DC.

Saaty, T. L. (1980). The analytic hierarchy process. New York: McGraw-Hill.

APPENDIX A

This list of strategic outcome metrics that has been proposed by the Defense Standards Improvement Council's Tiger Team.

Cost. Contractor overhead, cost premium for government unique requirements, consumable item price index, government administrative oversight, military specifications conversion price benefit.

Acquisition performance. Stability, consumable on-time deliveries, contract protests, contract defaults, supplier survey, contract changes.

Schedule. Acquisition phase time, administrative lead time, production lead time, multiyear procurements, FACNET transactions, logistics response time.

Commercial practices. Commercial contracts, cost and pricing data reduction, commercial content, contract specifications, commercial market share, credit card purchases.

APPENDIX B

COST

Consumable item price index. Total cost to DLA customers for representative sampling of consumable items expressed in constant dollars (FY$).

Military specification conversion price benefit. Cumulative cost avoidance using commercial specifications (performance specs, define (NGS), commercial item description (CIDs) and percent cost avoidance each year for Defense Logistics Agency (DLA) order. Applies only to items converted to commercial specifications since previous order.

ACQUISITION PERFORMANCE

Contract defaults. The total number of contract action defaults divided by the total number of contract actions.

Contract changes. Waivers and deviations of major significance and Class I Engineering Change Proposals (ECPs).

SCHEDULE

Acquisition phase time. Time for a system or item to progress from concept exploration and definition to start of the production and deployment phase. The metric is the average number of days between milestones (e.g., number of days between the signatures of MS 0 and MS I).

Administrative lead time. The average time from the signed formal requirements document to contract award. This excludes contracts for services and base support.

Multiyear procurements. The number and dollar value of multiyear procurements.

Federal Acquisition Computer Network (FACNET) transactions. The dollar amount of FACNET transactions divided by the total dollar amount of transactions in which simplified acquisition procedures have been used; the number of FACNET transactions divided by total number of transactions in which simplified acquisition procedures have been used.

Logistics response time. The time between customer order and customer receipt of DLA items.

COMMERCIAL PRACTICES

Contract specifications. The number of specifications and standards placed on contracts stratified by type: performance specifications (military performance [MIL-PRF], CIDs, Product Unique); NGS; old MIL-specs (assumed to be detailed); and, MIL-stds.

Credit card purchases. The number and dollar value of credit card purchases.

APPENDIX B

WHAT \ HOW	COST	ACQUISITION PERFORMANCE	SCHEDULE	COMMERCIAL PRACTICES	SYSTEM PERFORMANCE	TECHNOLOGY INNOVATION	
BETTER EW SYSTEMS	3	3	1	1	9	9	.40
CHEAPER EW SYSTEMS	9	3	3	3	1	1	.35
FASTER FIELDED EW SYSTEMS	9	3	3	3	1	3	.25
	.29	.13	.10	.10	.18	.21	

WHAT \ HOW	CONSUMABLE ITEM PRICE INDEX	MILITARY SPECS CONVER. ADVANT.	CONTR. DEFAULTS	CONTR. CHANGES	ACQUISITION PHASE TIME	ADMINISTRATIVE LEAD TIME	MULTI-YEAR PROCUREMENTS	FACNET TRANSACTIONS	LOGISTIC RESPONSE TIME	CONTR. SPECIFICATIONS	OVERHEAD	ACHIEVE OPER. GOALS	MTBF	INTEGRABILITY	
COST	9	9	3	9	9	3	9	3	3	9	9	3	3	3	.29
ACQUISITION PERFORMANCE	1	3	9	9	9	3	9	9	3	9	1	1	1	1	.13
SCHEDULE	1	1	3	9	9	3	9	3	1	1	1	1	1	1	.10
COMMERCIAL PRACTICES	1	9	1	1	1	1	1	3	1	1	1	1	1	1	.10
SYSTEM PERFORMANCE	1	3	1	3	1	1	1	1	1	3	1	9	9	3	.18
TECHNOLOGY INNOVATION	1	1	1	3	1	1	3	1	1	1	1	1	3	3	.21
	.06	.09	.05	.12	.10	.04	.11	.06	.04	.09	.06	.06	.07	.05	

WHAT \ HOW	PROGRAM OFFICE OVERHEAD	CONTRACTOR OVERHEAD	FUNDING STABILITY OVER TIME	UNIT PRODUCTION COST	UNIT LIFE-CYCLE COST	MIL. SPECS AND STANDS. ELIMINATED	MIL. SPECS & STANDS. REPLACED BY CIV.	COST OF PERFORMANCE			
CONSUMABLE ITEM PRICE INDEX											.06
MILITARY SPECS CONVERSION ADVANTAGE											.09
CONTRACT DEFAULTS											.05
CONTRACT CHANGES											.12
ACQUISITION PHASE TIME											.10
ADMINISTRATIVE LEAD TIME											.04
MULTIYEAR PROCUREMENTS											.11
FACNET TRANSACTIONS											.06
LOGISTIC RESPONSE TIME											.04
CONTRACT SPECIFICATIONS											.09
OVERHEAD											.06
ACHIEVE OPERATIONAL GOALS											.06
MTBF											.07
INTEGRABILITY											.05

WHAT \ HOW										
PROGRAM OFFICE OVERHEAD										
CONTRACTOR OVERHEAD										
FUNDING STABILITY OVERTIME										
UNIT PRODUCTION COST										
UNIT LIFE-CYCLE COST										
MIL. SPECS. AND STANDS. ELIMINATED										
MIL SPECS AND STNADS. REPLACE BY DIV.										
COST OF PERFORMANCE										