3-11-2011

Nuclear Enterprise Performance Measurement

Andrew S. Hackleman

NUCLEAR ENTERPRISE PERFORMANCE MEASUREMENT

THESIS

Andrew S. Hackleman, Major, USAF

AFIT-LSCM-ENS-11-05

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT-LSCM-ENS-11-05

NUCLEAR ENTERPRISE PERFORMANCE MEASUREMENT

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Logistics and Supply Chain Management

Andrew S. Hackleman

Major, USAF

March 2011

AFIT-LSCM-ENS-11-05

NUCLEAR ENTERPRISE PERFORMANCE MEASUREMENT

Andrew S. Hackleman,
Major, USAF

Approved:

//Signed//	16 March 2011
Dr. Alan W. Johnson (Chair)	Date
//Signed//	16 March 2011
LTC Darryl K. Ahner, Ph.D. (Member)	Date

AFIT-LSCM-ENS-11-05

Abstract

The criticality of the United States Air Force nuclear enterprise demands that commanders have the best possible understanding of system performance, both in the aggregate and at the drill-down levels sufficient to make timely corrective actions when warranted. We model a strategy-linked measurement system for nuclear enterprise sustainment. We propose a new Aggregation *h* method for aggregating performance metrics using United States Air Force approved or adapted metrics that possess the capability to weight metrics, as well as compare performance between organizations and within the same organization over time. We demonstrate our method with generated performance data designed to test the sensitivity of our method. Our Aggregation *h* method provides a simple, intuitive measurement approach that enables unity of effort and influences behavior at each hierarchical level towards achieving strategic goals, and is extendable to performance measurement for other complex sustainment systems.

AFIT-LSCM-ENS-11-05

This thesis is dedicated to my family for their patience and understanding throughout my time at the Air Force Institute of Technology.

Acknowledgments

I would like to offer sincere thanks to my family for their patient understanding throughout the AFIT experience, but especially for the many sequestered hours spent researching and writing this thesis. Also, I must acknowledge the keen insight and step-by-step guidance my adviser, Dr. Alan Johnson, gave me. Likewise, I would like to recognize my thesis reader, Dr. Darryl Ahner, for his frequent and pertinent advice. Finally, I would like to thank leaders at the Air Staff and Air Force Material Command who provided guidance and my sponsors Mr. Gregory Gross and Lt Col Ken Bottari of the Air Force Nuclear Weapons Center for the research topic and the complete academic freedom to allow the research to evolve into this thesis.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Acronyms

AHP:	Analytic Hierarchy Process
AWM:	Awaiting Maintenance
AWP:	Awaiting Parts
DoD:	Department of Defense
DoE:	Department of Energy
MCL:	Maintenance Capability Letter
MSE:	Maintenance Scheduling Effectiveness
MX:	Maintenance
NWRM:	Nuclear Weapons Related Material
PRP:	Personnel Reliability Program
RO:	Retrofit Order
SE:	Sustaining Engineering
TCTO:	Time-compliance Technical Order
UR:	Unsatisfactory Report
VFT:	Value-focused Thinking
WSA:	Weapons Storage Area

I. Introduction

Overview

This paper discusses United States Air Force nuclear enterprise performance measurement. The United States Air Force nuclear enterprise has come under fire in recent years for an unauthorized movement of warheads and an incorrect shipment of nuclear fuzes to Taiwan (Office of Secretary of Defense, 2008). As a result, it has had changes in leadership and organizational priorities and goals.

Nuclear weapons are a key part of the United States National Security Strategy (National Security Strategy, 2010). Nuclear weapons have a deterrent effect on the actions of other nations. In order for the United States to exercise the deterrent power of nuclear weapons, the deterrent must be credible. The Department of Energy and Department of Defense work together to maintain credible deterrence by ensuring the nation's nuclear stockpile is safe, secure, reliable and ready. The United States Air Force has custody of Department of Energy nuclear weapons and is charged with maintaining them in a state of readiness. The United States Air Force's obligation to the nation with regard to the sustainment of the nuclear stockpile is to enforce strict adherence to policy and technical guidance, which is integral to guaranteeing a safe, secure, reliable and ready nuclear stockpile.

The United States Air Force Chief of Staff, General Norton Schwartz, has made the nuclear enterprise the United States Air Force's number one priority (Nuclear Logistics Surety Implementation Plan, 2009). Spurred by recent high-profile incidents, the United States Air Force nuclear enterprise has come under tremendous internal and external scrutiny. The result of this scrutiny has been the identification of a large number of deficient and neglected areas. To address the deficiency and neglect, the United States Air Force has undertaken an aggressive campaign to reinvigorate the nuclear enterprise and has taken a number of meaningful steps to do so, beginning in 2007 (Nuclear Logistics Surety Implementation Plan, 2009).

Background

United States Air Force logistics leadership has developed a method to track and oversee the campaign to improve (or reinvigorate) the sustainment of the nuclear enterprise. They have created 15 outcome areas that allow categorization of ongoing improvement areas that span the sustainment mission in the nuclear enterprise. These outcomes are reviewed by United States Air Force leaders. In terms of performance measurement, this set of outcomes is how the United States Air Force measures and monitors improvement in key areas of the sustainment of the nuclear enterprise (Nuclear Logistics Surety Implementation Plan, 2009).

The United States Air Force nuclear enterprise faces many challenges. Perhaps the most resource- and time-consuming are those challenges stemming from efforts to address findings from several reports--Scheslinger Report, Admiral Kirkland Donald Report, United States Air Force Blue Ribbon Review, Defense Science Board, Minot Commander Directed Investigation--which includes gaining accountability for nuclear weapons related material, deconflicting Department of Energy, Department of Defense and United States Air Force policy, standardizing the inspection process, to name a few. Not only does the United States Air Force have to manage ongoing external scrutiny, but it must also work diligently to make meaningful improvements the areas found to be deficient or neglected.

In addition to the challenges outlined, the United States Air Force nuclear enterprise must also contend with an aging nuclear stockpile and critical nuclear infrastructure in a scarce resource environment (a challenge shared by conventional United States Air Force weapons systems and infrastructure). In order to meet these challenges head-on, the United States Air Force will need to have clear strategic objectives and a means to measure performance that is directly linked to these objectives, from sustainment at the unit level to decision-makers at the Air Force Nuclear Weapons Center, Major Commands and Air Staff.

The United States Air Force currently measures performance in three ways: monitoring the improvement of deficient and neglected areas in the nuclear enterprise areas identified by the aforementioned reports, Status of Resources and Training System and through various, frequent inspections, which include United States Air Force and Department of Defense Nuclear Surety Inspections, Logistics Compliance and Assessment Program, Nuclear Operational Readiness Inspections and a few compliance oriented periodic internal assessments.

The first area of measurement is a rapidly evolving effort and has been directed at answering report findings and ensuring the United States Air Force has an adequate performance baseline moving forward. Starting about 2008 this was done by measuring a set of 15 desired outcomes, which were championed by Colonels (or equivalent) responsible for monitoring and measuring improvement in their outcome area (Nuclear Logistics Surety Implementation Plan, 2009). This type of measurement is relatively new to the nuclear enterprise and has been an important tool for shepherding the United States Air Force nuclear enterprise on the path towards reinvigorating the nuclear enterprise, but these measurements were not designed to measure organizational performance, based on a strategic objective. Rather, they are focused on specific, isolated outcomes. The current stage of evolution has the United States Air Force starting a transition from measuring the 15 desired outcomes—marking the end of reinvigoration and the beginning of continuing to strengthen the nuclear enterprise—to a system that consolidates the outcomes into four measured areas and a number of performance metrics identified to measure criteria in this fledgling performance measurement system (Maj Gen Close, 2010). Another nascent performance measurement system, drawing from the original Nuclear Logistics Surety Implementation Plan is being developed by a separate USAF headquarters office, based on the top-level criteria identified in the document. Although both these measurement systems have top-level strategic goals, neither uses a definition of sustainment consistent with USAF and DoD lifecycle management, which is the common approach for other USAF systems (DoDD 5000.01, 2003).

The second area, Status of Resources and Training System, measures the readiness of Designed Operational Capability. Status of Resources and Training System measures the capability of a unit to go to war; it does not measure sustainment performance (Air Force Instruction 10-201, 2006). The United States Air Force, Department of Defense and congress only see the non-negotiable performance floor via Status of Resources and Training System, so any variance from full capability related to nuclear enterprise sustainment will experience significant lag and indicate significant performance degradation. Finally, the United States Air Force relies on inspection data to measure performance in the nuclear enterprise. Indeed, inspection results do provide insight into compliance and, to a certain extent, performance. However, measuring performance through inspection has serious limitations, such as a small sample of data relative to total population of sustainment data, which makes trending and decision-making, with respect to sustainment performance, ineffective. That is not to say that inspection doesn't provide a good measure of compliance, it does. However, compliance should be viewed as one of many dimensions of performance (Eccles, 1991).

So, despite measuring improvement, capability and compliance, the United States Air Force nuclear enterprise sustainment lacks a strategy-linked system of performance measurement that can be meaningfully aggregated at decision-maker (or hierarchical) levels. A strategy-linked performance measurement system is crucial, because it positively influences behavior toward strategic goals and enables unity of effort at each hierarchical level (Neely, 1995).

The United States Air Force recognizes the lack of nuclear enterprise performance measurement and is working to develop sustainment performance metrics as it transitions from monitoring 15 outcomes and answering findings from various reports (Maj Gen Close, 2010). The goal of this paper is to contribute to United States Air Force efforts and influence the development of a performance measurement system, particularly with regard to a performance measurement hierarchy and a method for aggregating metrics within the hierarchy. Establishing such a system is essential to achieving the strategic sustainment goal, because measuring influences behavior and enables unity of effort (Neely, 1995). As the United States Air Force begins to take action to develop a performance measurement system, it is crucial that these measurements be designed based on strategic goals and linked through a meaningful system of aggregation. This will ensure that the metrics are measuring the right things, from a strategic perspective.

This paper explores the lack of a performance measurement system in the United States Air Force and discusses why and how performance measurement should be designed for the United States Air Force nuclear enterprise. The importance of performance measurement is outlined and an overview of the United States Air Force nuclear enterprise and its current state is presented, followed by a discussion of the challenges facing the nuclear enterprise and lack of a performance measurement system. Finally, a model of a strategy-linked performance measurement system is presented, demonstrating a technique for aggregating performance measurements at decision-maker, or hierarchical, levels.

Motivation

The original vector for our research was to determine whether the United States Air Force nuclear enterprise is effectively managing time compliance technical orders. The follow-on to this topic was to answer the question: how do we know time compliance technical orders are or are not being effectively managed? We quickly determined that the United States Air Force doesn't measure time compliance technical order management. Additionally, because the United States Air Force nuclear enterprise maintains both United States Air Force and Department of Energy items, for which time compliance technical orders and retrofit orders performed, different process and policies applied.

In order to determine if the United States Air Force nuclear enterprise effectively manages time compliance technical order and answer the question, "how do we know?" we knew that we would need historical data that is not currently analyzed and, indeed, may not even be collected. Simply stated, there are sustaining engineering, field maintenance and supply aspects to measuring effective management of time compliance technical order (and by extension, retrofit orders). To take an enterprise view of the management of time compliance technical orders, we are really concerned with the process of configuration management, under which the development, funding and execution of time compliance technical orders and retrofit orders would fall.

Understanding what would be required to study the effectiveness of United States Air Force nuclear enterprise configuration management orders led us to the broader awareness that the nuclear enterprise lacks a coherent, strategy-linked performance measurement system. In such a system, presumably, nuclear enterprise configuration management would figure prominently.

So, motivated by our initial challenge to measure configuration, we determined that creating the framework of performance measurement for nuclear enterprise sustainment was a necessary first step and would provide the context and understanding of how and where configuration management fits into sustaining the nuclear enterprise. Although there are ongoing efforts to design a method for measuring nuclear enterprise sustainment performance, the United States Air Force nuclear enterprise lacks a strategy-linked performance measurement system. We focused on developing a performance measurement hierarchy with nuclear enterprise sustainment as the strategic goal.

A performance measurement system will allow leaders at all levels to accurately assess the health of nuclear enterprise sustainment and help inform the Planning, Programming, Budgeting and Execution (PPBE) process (Haines, 2009).

II. Literature Review

Performance Measurement

Performance measurement is a topic for which there has been a great deal of academic research. However, performance measurement also has potential for misapplication in organizations. The literature agrees that performance measurement must be designed with the organization's strategic goals as the centerpiece, and that a direct link should be made between strategic goals and the organizational business processes that produce outputs that achieve strategic goals, but organizations often stray from academic guidelines (Neely, 1995). Therefore, strategic goals should be measured as a composite of key outputs that inform leadership about the performance of the organization. The top level composite measure of the organizations strategic goal should be capable of disaggregating and cascading down through the organization to key outputs that can be directly measured. By establishing this strategic linkage, the organization can be assured that there is a functional relationship between the lower level output measurements and the strategic goal. Additionally, a strategic linkage of performance measures ensures the organization is measuring the *right* outputs and prevents measuring too much (Brignall, 2000). If an organization doesn't develop a performance measurement system based on strategic goals, it runs the risk of measuring too much and the wrong outputs. Further, without a strategic linkage, managers at all levels within the organization will not be able to benefit from the positive side of performance measurement: influencing behavior. When performance measurements are linked to the organization's strategic goals and aggregated at appropriate management levels, they will influence behavior to achieve organizational goals. Performance measurements that are not based on organizational strategy will also influence behavior, but this behavior may not necessarily be aligned with organizational strategy, and the measurements may even conflict with one another (Brignall, 2000).

All large organizations measure performance (Brignall, 2000). In order to remain viable and competitive, organizations must measure performance. Of course, performance measurement has pitfalls that can actually damage an organization as much as not measuring performance at all. These pitfalls occur when organizations measure too much or the wrong outputs (Brignall, 2000). If an organization is lost in the minutia of a large number of meaningless measurements, managers will become bogged down by conflicting and unnecessary measures and the organization will not move toward its strategic goals (Gunasekeran, 2004). Likewise, when organizations choose to measure the wrong things, there is a misalignment between the performance measurements that managers use to make decisions and the strategic goals of the organization. Either of these measurement mistakes can cause an organization to underperform and fail to achieve strategic goals.

Quantitative measurement has power to influence behavior: positive or negative (Neely, 1995). As a result, performance measurement is crucial to achieving strategic organizational goals. However, the critical first step in measuring performance is determining how the system of measurement is to be developed. The process of building the system must start at the top with the strategic goal and be linked in a meaningful way to key outputs that measure the performance of the organization in key areas that contribute toward achieving strategic goals. Without this linkage, organizations are likely to suffer the pitfalls of performance measurement, as discussed in the previous paragraph.

There are two leading methods for developing a performance measurement system in academic literature: framework and process (Neely, 1995). The framework method uses a specific set of criteria for measuring performance. The process method outlines a number of steps to take in developing a strategically aligned performance measurement system, which, unsurprisingly lead to unique outcomes for each organization.

Perhaps the most well-known performance measurement framework method is The Balanced Scorecard. The Balanced Scorecard has gained popularity in business over the last decade. It takes four questions (criteria) and develops performance measures for each one. The areas below make up the "scorecard" and it is balanced because each of the four elements of the scorecard makes up some proportion of the total, which is 100 percent (Neely, 1995).

- How do we look to our shareholders (financial perspective)?
- What must we excel at (internal business perspective)?
- How do our customers see us (customer perspective)?
- How can we continue to improve and create value (innovation and learning perspective)?

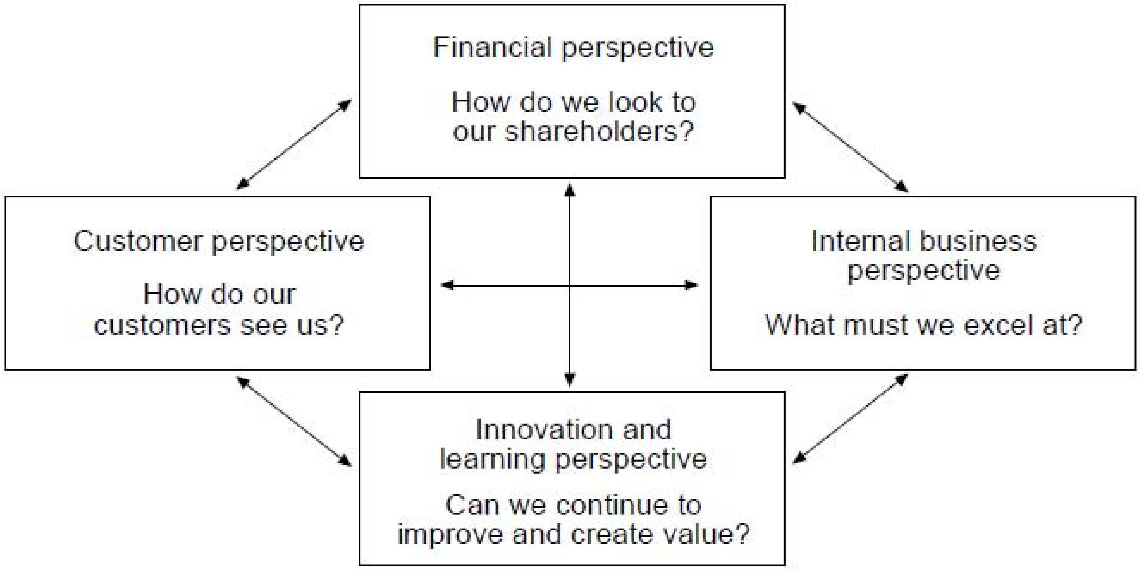


Figure 2-1. Balanced scorecard model.

The Balanced Scorecard has evolved since its initial rise to popularity. It focuses less on balance. That is, many successful users of this method find that balance is not necessarily a good thing with respect to performance measurement. For example, it may often be advisable to tip the balance of the scorecard to focus on the customer perspective. A criticism of The Balanced Scorecard is that it doesn't explicitly take into account the performance of other like organizations (i.e. competition) and its criteria, which are foundations of the method, may be arbitrary and not fit some organization (Centre for Business Performance, 2004).

The other method of performance measurement design uses a process instead of framework to develop a unique, strategically aligned system of performance measurement. The process method, like The Balanced Scorecard method, asks a series of questions to determine an organization's strategic goals and objectives and how to measure them. However, unlike The Balanced Scorecard, the resulting system of measurement isn't bound by maintaining a balance (the organization decides how important each area is) or fitting measurements into four prescribed categories, which for some organizations could arbitrary. For the United States Air Force, the four balanced scorecard measurement areas do not directly translate into analogs in government organizations, so any attempt to translate these areas would be subjective at best and arbitrary (without meaning) at worst.

The process method removes the need to wrestle measurement areas into arbitrary categories, but follows the spirit of performance measurement theory, which universally agrees that measurement needs to be aligned with strategy, as the effect of measuring is the stimulation of action (Neely, 1995). The action stimulated is either toward the organization's strategic goal or it isn't. In other words, if the actions of subordinate organizations aren't measuring performance in a way that directly supports strategic goals, their efforts will act like dead weight or even work against organizational strategy.

The following captures the essential elements of using the process method of performance measurement system design (Neely, 1995):

- Performance criteria must be chosen from the company's objectives.
- Performance criteria must make possible the comparison of organizations which are in the same business.
- The purpose of each performance criterion must be clear.
- Data collection and methods of calculating the performance criterion must be clearly defined.
- Ratio-based performance criteria are preferred to absolute number.
- Performance criteria should be under control of the evaluated organizational unit.
- Performance criteria should be selected through discussions with the people involved

(customers, employees, managers).

- Objective performance criteria are preferable to subjective ones.

It's easy to see the utility of the process method and the flexibility it allows organizations, such as the United States Air Force, that aren't organized like a typical U.S. corporation.

Aggregation

Analytic Hierarchy Process

Analytic Hierarchy Process is a multicriteria decision-making system (Saaty, 1990). Analytic Hierarchy Process has gained popularity in a variety of fields requiring complex, multicriteria decision-making. The process breaks down complex decisions, or goals, into a hierarchy of constituent parts. These parts are prioritized by a decision-maker and a pairwise comparison is made. The Analytic Hierarchy Process breaks down the goal of the organization, which is a complex problem that a decision-maker doesn't have control or direct influence over, into smaller, more general criteria that directly relate to the overall goal or problem and which the decision-maker can control. The process of building the hierarchy is carried out until the goal is broken down into the smallest possible, while still meaningful, sub-criteria. "The basic principle to follow in creating this structure is always to see if one can answer the following question: Can I compare the elements on a lower level using some or all of the elements on the next higher level as criteria or attributes of the lower level elements?" (Saaty, 1990). In a 1990 article, Thomas L. Saaty outlined a 10-step process for constructing the hierarchy (Saaty, 1990):

1. Identify the overall goal. What are we trying to accomplish? What is the main question?

2. Identify the subgoals of the overall goal. If relevant, identify time horizons that affect the decision.
3. Identify criteria that must be satisfied to fulfill the subgoals of the overall goal.
4. Identify subcriteria under each criterion. Note that criteria or subcriteria may be specified in terms of ranges of values of parameters or in terms of verbal intensities such as high, medium, low.
5. Identify the actors involved.
6. Identify the actors' goals.
7. Identify the actors' policies.
8. Identify options or outcomes.
9. For yes-no decisions, take the most preferred outcome and compare the benefits and costs of making the decision with those of not making it.
10. Do a benefit/cost analysis using marginal values. Because we are dealing with dominance hierarchies, ask which alternative yields the greatest benefit; for costs, which alternative costs the most, and for risks, which alternative is more risky.

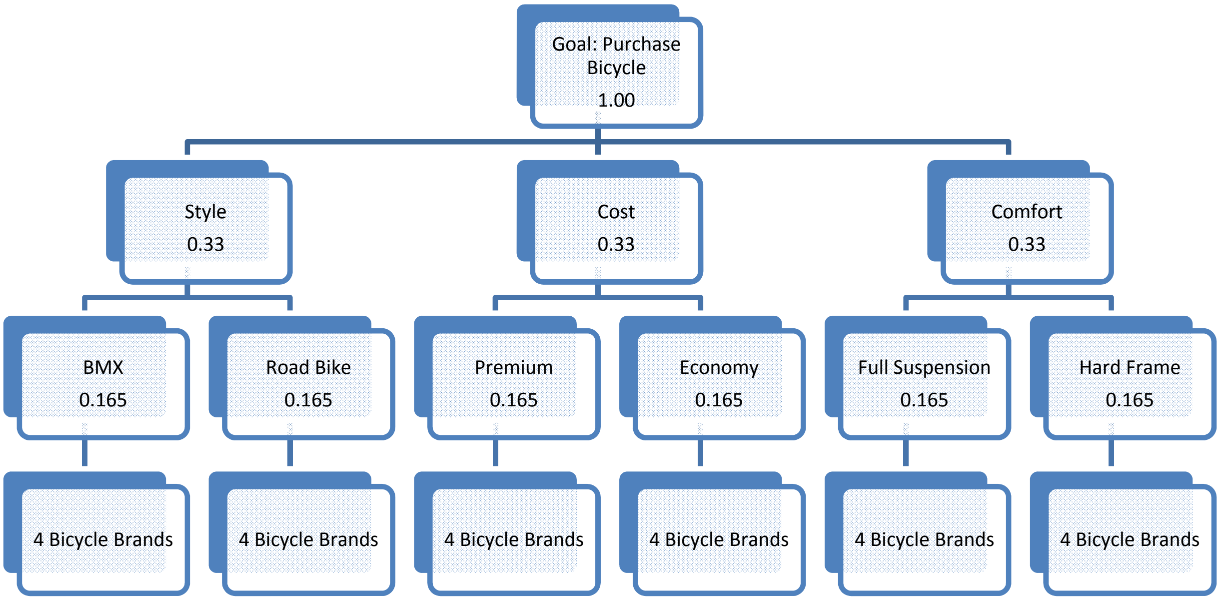


Figure 2-2. Analytic Hierarchy Process bicycle purchase example.

A simple example depicted in Figure 2-2 shows an Analytic Hierarchy Process model for buying a bicycle. The process starts by identifying the goal (in this case buying a bicycle), which takes on a priority value of 1.00. The first set of criteria is called general. General criteria break down into secondary subcriteria, tertiary criteria and so on. For this example, only general and secondary subcriteria are used.

Each subcriterion is given a weight, as judged by a decision-maker. The weighting system for Analytic Hierarchy Process is defined as follows:

1	Equal Importance
3	Moderate Importance
5	Strong Importance
7	Very Strong Importance
9	Extreme Importance
2,4,6,8	Compromise values
Reciprocals of above values	above nonzero numbers assigned to it when compared to activity j, then
Rationals	Ratios arising from the scale
1.1-1.9	Used for tied activities

Table 2-1. Analytic Hierarchy Process priority scheme.

Once the alternatives are given a weight, a pairwise comparison of the criteria is done in a square matrix. The resulting ratios now make up a matrix. The matrix is now squared and the sum of each row is divided by the sum of the matrix, giving the eigenvector, which normalizes the comparisons. The matrix is squared again until the difference between the eigenvectors is minimized to a predetermined significant digit (usually four decimal places) (Saaty, 1990). Now that the criteria priorities are determined via eigenvectors, the same process is applied to the alternatives; in this case the four bicycle choices. These comparisons can be made in terms of subjective judgments or subjective scoring as outlined above, but the comparisons can also be made on the basis of quantitative measures, providing the units and scale are the same (Johnson, 2007). For example, cost can be quantitatively measured, by taking the sum of the sum of the total cost of our bicycles and dividing each bicycle cost by the total. This normalizes the cost in terms of a ratio of the each brand to the total. Now, to complete the Analytic Hierarchy Process, all that remains is to multiply the eigenvector values for each alternative against the eigenvectors for the decision criteria. The result is a one column, four row matrix with a score based on normalized values for decision criteria and alternatives. The alternative with the highest value, based on pairwise comparisons at each level of the decision hierarchy, is the alternative that best matches the criteria to achieve the goal.

Analytic Hierarchy Process is a powerful tool for making multi-criteria decision by breaking down the overall goal into smaller and smaller constituent parts, where the smaller constituent parts represent criteria that can be controlled and quantified (or at least qualitatively judged). And, logically, by determining priority for these constituent parts the alternative with the largest eigenvector for each subcriterion up through the hierarchy will be selected as the best alternative; one that best accomplished the top level goal.

It is the hierarchy and aggregation aspects of Analytic Hierarchy Process that make it a good method for making sense of metrics in the nuclear enterprise. As long as the lower level metrics are standardized and a decision-maker prioritizes the subcriteria, the aggregation is meaningful, in terms of a top level metric. In other words, if instead of purchasing a bicycle we were trying to determine the overall performance of an organization, Analytic Hierarchy Process can be used to determine how well subordinate units and business processes are performing with respect to achieving the overall goal (Johnson, 2007). For this research, the overall goal is nuclear enterprise sustainment.

Value-Focused Thinking

Value-focused thinking (VFT) is a way of approaching multi-criteria decision analysis. VFT has three major tenets: start with values, generated better alternatives and use the values started with to evaluate the alternatives (Parnell, 2008). The values stated with are the decision-maker's goals. The values are used to generate acceptable alternatives, given the decision-maker values. Once a spectrum of alternatives has been identified, the values are used in an appropriate multi-objective decision analysis.

VFT is also used to make qualitative value model. The qualitative value modeling is a four step process: 1) identify fundamental objective; 2) identify functions that provide value; 3) identify objectives that define value; and 4) identify value measures (Parnell, 2008).

Step 1 requires the analyst to identify the fundamental, or strategic, objective. The fundamental objective must be clearly defined and understood. It is essential that the objective be understood by stakeholders, because the alternative selection ultimately relies on the fundamental objective.

Step 2 is to identify functions that provide value to the fundamental objective. In this step, all of the key processes, functions or relationships are identified that contribute value to the fundamental objective.

Step 3 is to identify the functions that provide value. This step determines the objectives that define value for the fundamental objective. This step may result in identifying sub-objective to the fundamental objective, followed by the identification of value measures.

Step 4 is identifying value measures. Value measures can be identified by research, interviews with subject-matter experts, and decision makers (Parnell, 2008). Above all, value measures must be aligned with the objective. The alignment may be either direct or by proxy. The direct measure directly measures the objective. A proxy measure focuses on a parallel process that is closely correlated with the objective.

VFT uses multiple objective decision analysis to select alternative in the value model. One simple method is the additive value model. It uses the simple additive equation in Equation 1 to determine each alternative's value.

$$v(x) = \sum_i^n w_i \, v_i(x_i) \qquad (1)$$

Where v(x) is the alternatives value

i=1 to n is the number of the value measure

x_i is the alternatives score on the ith measure

$v_i(x_i)$= is the single dimensional value of a score of x_i

w_i is the weight of the i^{th} measure shown in Equation 2 where:

$$\sum_i^n w_i = 1 \qquad (2)$$

Aggregation Metric D

Another method of aggregation, not currently used in logistics applications, is a variant of the geometric mean. The geometric mean is used in aggregation applications in biological science, economic indices, and finance. The properties of the geometric mean make it well suited for aggregating performance metrics. We chose to pursue the geometric mean and borrowed techniques from economic indexing and environmental sustainability aggregation techniques. The algorithm used in this research is discussed in detail in the methodology chapter.

Aggregate metric D is a method developed to aggregate environmental sustainability metrics (Sikdar, 2009). It is used by the Environmental Protection Agency to help determine which biofuels are most sustainable. The method uses a variation of the geometric mean. It takes the product of a vector of ratios x_i/y_i, where x_i is the state of a system S_1 $(x_1, x_2, \ldots, x_n)$ and y_i is the state of a system S_2 $(y_1, y_2, \ldots, y_n)$, to the nth root. A linear weight c_i can also be applied to the aggregation, as shown in Equation 3 (Sikdar, 2009).

$$D = [\prod_{i=1}^{n} c_i (y_i / x_i)]^{1/n} \tag{3}$$

This method is simple, but effective at making system comparisons over time. Also, because of the properties of the geometric mean, the central tendency of the systems will be accurately calculated.

We considered using this method and tested a model using the algorithm, but determined that it wasn't suited to logistics aggregation, because the Aggregate Metric D compares system states one month to the next. Directly applied to logistics applications, the aggregation method will return deceptive values. For example, using this method as designed, if we compare the same metric of two organizations, the Aggregate Metric D will compare each organization's performance at two different states (i.e. current month compared to previous month). This comparison will provide an accurate report of the relative performance of the organization from one month to the next, but it doesn't enable a meaningful comparison between the two organizations, because even if the organizations are performing differently, the comparison month to month will only compare the organizations previous month's performance. We illustrate a simple example in Table 2-2 that assumes a comparison between two similar organizations, where good performance is indicated by a higher percentage value. The illustration shows that despite an obvious difference in performance, the poor performing organization X actually reports a higher Aggregate Metric D value. Using the Aggregate metric D, as designed, we would rank the poor performing organization higher than the good performing organization, due to the comparison to system states relative only to each organization's previous month's performance. Nuclear enterprise sustainment performance requires an aggregation method that closely represents the constituent metric values.

	Aggregate Metric D		
	Month 1	Month 2	Aggregate Value
Organization X Performance	55%	56%	102%
Organization Y Performance	100%	97%	97%

Table 2-2. Aggregate Metric D illustration.

In the Table 2-2 illustration, we do not apply the c_i weight, as a linear weight in a multiplicative model doesn't influence the geometric distance between the metric values, it only serves to scale the product. This is another factor in our decision to pursue an alternative aggregation method, as we require the ability to differentiate between the importance and influence of individual metrics.

Definition of Strategic Goal--Sustainment

The first step in creating a performance measurement hierarchy for nuclear enterprise sustainment was to carefully define the meaning of sustainment. We based the construction of the sustainment performance measurement hierarchy on the definition and description of sustainment found in Defense Acquisition Guidebook, Department of Defense Instruction 5000.02, Operation of the Defense Acquisition System, and Department of Defense Directive 5000.01, The Defense Acquisition System. According to paragraph 3.9.2.1., the Defense Acquisition Guidebook defines sustainment as follows (Defense Acquisition Guidebook, 2010):

> Sustainment includes supply, maintenance, transportation, sustaining engineering, data management, configuration management, manpower, personnel, training, habitability, survivability, environment, safety (including explosives safety), occupational health, protection of critical program information, anti-tamper provisions, and information technology (IT), including National Security Systems (NSS), supportability and interoperability functions." In addition, according to paragraph 5.4.3 (Sustainment: Operations and Support), "while acquisition phase activities are critical to designing and implementing a successful and affordable sustainment strategy, the ultimate measure of success is application of that strategy after the system has been deployed for operational use. Total Life Cycle Systems Management, through single point accountability, and Performance Based Logistics, by designating performance outcomes vs. segmented functional support, enables that objective. Warfighters require operational readiness and operation effectiveness - systems accomplishing their missions in accordance with their design parameters in a mission environment. Systems, regardless of the application of design for supportability, will suffer varying stresses during actual operational deployment and use.

The Department of Defense Directive 5000.01 definition states (DoD Directive 5000.01, 2003):

> Sustainment involves the supportability of fielded systems and their subsequent life cycle product support - from initial procurement to supply chain management (including maintenance) to reutilization and disposal. It includes sustainment functions such as initial provisioning, cataloging, inventory management and

warehousing, and depot and field level maintenance. Sustainment begins when any portion of the production quantity has been fielded for operational use. Sustainment includes assessment, execution and oversight of performance based logistics initiatives, including management of performance agreements with force and support providers; oversight of implementation of support systems integration strategies; application of diagnostics, prognostics, and other condition based maintenance techniques; coordination of logistics information technology and other enterprise integration efforts; implementation of logistics footprint reduction strategies; coordination of mission area integration; identification of technology insertion opportunities; identification of operations and support cost reduction opportunities and monitoring of key support metrics.

Adding to the definitions in the Department of Defense guidance, "Designing and Assessing Supportability in Department of Defense Weapon Systems: A Guide to Increased Reliability and Reduced Logistics Footprint provides detailed instruction for system acquisition and lifecycle management", released in 2003, provides a great deal of insight into how the sustainment phase of lifecycle management should be viewed. In particular, the guide makes an explicit link between performance and sustainment (as can be inferred from the sustainment definitions), where performance (i.e. reliability, maintainability, availability and process efficiency) is a measure of sustainment Operations and Support investment. In other words, system performance is a function of investment in lifecycle sustainment (Haines, 2009). Thus performance is the key measure of sustainment (Office of Secretary of Defense, 2003).

The Department of Defense also provides a detailed description of Program Manager responsibilities. The Program Manager is responsible for the weapon system for the entire lifecycle, including sustainment (DoD Directive 5000.01, 2003). As mentioned above, the Department of Defense Directive, Department of Defense Instruction and guide emphasize the importance of sustainment and articulate an explicit link between sustainment and performance, the latter being a function of the former (Office of Secretary of Defense, 2003). According to the Department of Defense, sustainment encompasses a range of performance areas, illustrated by figure 4, where System Operational Effectiveness is the overall goal of sustainment (Office of Secretary of Defense, 2003). System Operational Effectiveness is defined by Technical Effectiveness and Process Efficiency. Within the Technical Effectiveness category is System Performance, which is determined during pre-acquisition and acquisition, and System availability.

Combined with the expansive definition of sustainment, as detailed by the Department of Defense, we drew heavily from key leaders within the nuclear enterprise. Our approach was to ask nuclear enterprise leaders what they believed was important to measure, discuss with them the Department of Defense sustainment definition and show them a working model of the performance measurement hierarchy. This was an iterative process that involved leaders at all levels of the nuclear enterprise, which included senior noncommissioned officers, civilians and officers up to the rank of Major General. Interestingly, there was no significant difference of opinion, despite interviewing more than a dozen leaders.

III. Article Manuscript

Nuclear Enterprise Performance Measurement

Andrew S. Hackleman, Alan Johnson and Darryl K. Ahner
Air Force Institute of Technology
Wright-Patterson Air Force Base, Ohio

Abstract

The criticality of the United States Air Force nuclear enterprise demands that commanders have the best possible understanding of system performance, both in the aggregate and at the drill-down levels sufficient to make timely corrective actions when warranted. We model a strategy-linked measurement system for nuclear enterprise sustainment. We propose a new Aggregation *h* method for aggregating performance metrics using United States Air Force approved or adapted metrics that possess the capability to weight metrics, as well as compare performance between organizations and within the same organization over time. We demonstrate our method with generated performance data designed to test the sensitivity of our method. Our Aggregation *h* method provides a simple, intuitive measurement approach that enables unity of effort and influences behavior at each hierarchical level towards achieving strategic goals, and is extendable to performance measurement for other complex sustainment systems.

Keywords

Performance measurement, process measurement, strategy, multicriteria decision-making, aggregation

1. Introduction

Nuclear weapons are a key part of the United States National Security Strategy (National Security Strategy, 2010). Nuclear weapons have a deterrent effect on the actions of other nations. In order for the United States to exercise the deterrent power of nuclear weapons, the deterrent must be credible. The Department of Energy and Department of Defense work together to maintain credible deterrence by ensuring the nation's nuclear stockpile is safe, secure, reliable and ready. The United States Air Force has custody of Department of Energy nuclear weapons and is charged with maintaining them in a state of readiness. The United States Air Force's obligation to the nation with regard to the sustainment of the nuclear stockpile is to enforce strict adherence to policy and technical guidance, which is integral to guaranteeing a safe, secure, reliable and ready nuclear stockpile.

Despite rigorous and frequent inspections, the United States Air Force nuclear enterprise sustainment lacks a strategy-linked system of performance measurement that can be meaningfully aggregated at decision-maker (or hierarchical) levels. The United States Air Force recognizes the lack of nuclear enterprise performance measurement and is working to develop sustainment performance metrics as it transitions from monitoring 15 outcomes, instituted to reinvigorate the nuclear enterprise, and answering findings from various reports (Maj Gen Close, 2010). The goal of this paper is to contribute to United States Air Force efforts and influence the development of a performance measurement system; specifically a performance measurement hierarchy and a method for aggregating metrics within the hierarchy. Establishing such a system is essential to achieving the strategic sustainment goal, because measuring influences behavior and enables unity of effort (Neely, 1995). As the United States Air Force begins to take action to develop a performance measurement system, it is crucial that these measurements be designed based on strategic goals and linked through a meaningful system of aggregation. This will ensure that metrics are measuring the right things, from a strategic perspective.

This paper explores the lack of a performance measurement system in the United States Air Force and discusses why and how performance measurement should be designed for the United States Air Force nuclear enterprise. The importance of performance measurement is outlined and an overview of the United States Air Force nuclear enterprise and its current state is presented. We then introduce a strategy-linked performance measurement system model, and demonstrate a technique for aggregating performance measurements at decision-maker, or hierarchical, levels.

1.2 Performance Measurement

Performance measurement is a topic for which there has been a great deal of academic research. Despite this, however, performance measurement also has potential for misapplication in organizations. The literature agrees that performance measurement must be designed with the organization's strategic goals as the central focus, and that a direct link should be made between strategic goals and the organization's business processes that produce outputs that achieve strategic goals (Neely, 1995). When performance measurements are linked to the organization's strategic goals and aggregated at appropriate management levels, they will influence behavior to achieve organizational goals (Brignall, 2000).

Performance measurement, done badly, can damage an organization. These pitfalls happen when organizations either attempt to measure too much or measure the wrong outputs. If an organization becomes lost in the minutia of a large number of measurements, managers can become bogged down by conflicting and unnecessary measures and the organization will not move toward its strategic goals (Brignall, 2000). Likewise, when organizations choose to measure the wrong things, there is a misalignment between the performance measurements that managers use to make decisions and the organization's strategic goals. Either of these measurement mistakes can cause an organization to underperform and fail to achieve strategic goals.

There are two leading methods for developing a performance measurement system in academic literature: framework and process (Neely, 1995). The framework method uses a specific set of criteria for measuring performance. Conversely, the process method outlines a number of steps to take in developing a strategically aligned performance measurement system, which can lead to unique outcomes for each organization.

The process method removes the need to wrestle measurement areas into arbitrary categories, but follows the spirit of performance measurement theory, which universally agrees that measurement needs to be aligned with strategy, as the effect of measuring is the stimulation of action (Neely, 1995). The action stimulated is either toward the organization's strategic goal or it isn't. The following captures the essential elements of using the process method of performance measurement system design (Neely, 1995):

- Performance criteria must be chosen from the company's objectives.
- Performance criteria must make possible the comparison of organizations which are in the same business.
- The purpose of each performance criterion must be clear.
- Data collection and methods of calculating the performance criterion must be clearly defined.

- Ratio-based performance criteria are preferred to absolute numbers.

- Performance criteria should be under control of the evaluated organizational unit.

- Performance criteria should be selected through discussions with the people involved (customers, employees, managers).

- Objective performance criteria are preferable to subjective ones.

The process of hierarchy construction starts with identifying the strategic goal. A set of subcriteria are then determined that, taken together, comprise the goal. The subcriteria may be further decomposed into tertiary subcriteria. Finally, outputs are identified for each subcriterion that meaningfully measure and collectively define the particular subcriterion they support.

Figure 3-1. Theoretical performance measurement hierarchy model

Constructing a performance measurement hierarchy is the first major step toward realizing a strategy-linked performance measurement system. The next step is to determine the simplest meaningful way to quantitatively link the criteria and metrics set forth in the performance measurement hierarchy. That is, how should lower level output metrics be aggregated at each successive hierarchical level? We review three candidate approaches: the Analytic Hierarchy Process, Value Focused Thinking, and variations of the geometric mean.

1.3 Aggregation

Analytic Hierarchy Process

Analytic Hierarchy Process (AHP) is a multicriteria decision-making system (Saaty, 1990). AHP has gained popularity in a variety of fields requiring complex, multicriteria decision-making. The process breaks down complex decisions, or goals, into a hierarchy of constituent parts. These parts are prioritized by a decision-maker and a pairwise comparison is made. AHP allocates the organization's goal, which may be a complex problem that a decision-maker doesn't have control or direct influence over, into smaller, more general criteria that both directly relate to the overall goal or problem and are under the decision-maker's control. The process of building the hierarchy is carried out until the goal is broken down into the smallest possible, while still meaningful, sub-criteria. "The basic principle to follow in creating this structure is always to see if one can answer the following question: Can I compare the elements on a lower level using some or all of the elements on the next higher level as criteria or attributes of the lower level elements?" (Saaty, 1990). By determining priority for these constituent parts the alternative with the largest eigenvector for each subcriterion up through the hierarchy will be selected as the best alternative; one that best accomplishes the top level goal.

Value-Focused Thinking

Value-focused thinking (VFT) represents another way of approaching multi-criteria decision analysis. VFT has three major tenets: identify starting values, generate acceptable decision alternatives and use the values started with to evaluate the alternatives (Parnell, 2008). The starting values are the decision-maker's goals. After a set of decision alternatives have been identified, the values are used in an appropriate multi-objective decision analysis.

VFT uses multiple objective decision analysis to rank alternatives in the value model. One simple ranking method is the additive value model shown in Equation 1:

$$v(x) = \sum_{i=1}^{n} w_i \, v_i(x_i) \tag{1}$$

where

$v(x)$ is a decision alternative's overall value

x_i is the alternative's score on the i^{th} measure for i = 1, …, n criteria

$v_i(x_i)$ is the single dimensional value of score x_i

w_i is the weight of the i^{th} measure shown in Equation 2 where:

$$\sum_{i=1}^{n} w_i = 1 \tag{2}$$

Aggregation Metric D

Another method of aggregation, not currently used in logistics applications, is a variant of the geometric mean. The geometric mean is used in aggregation applications in biological science, economic indices, and finance. The properties of the geometric mean make it well suited for aggregating performance metrics. Aggregation metric D is a method developed to aggregate environmental sustainability metrics (Sikdar, 2009). It is used by the Environmental Protection Agency to help determine which biofuels are most sustainable. The method uses a variation of the geometric mean. It takes the product of a vector of ratios x_i/y_i, where x_i is the state of a system S_1 $(x_1, x_2, \ldots, x_n)$ and y_i is the state of a system S_2 $(y_1, y_2, \ldots, y_n)$, to the n^{th} root. A linear weight c_i can also be applied to the aggregation, as shown in Equation 3 (Sikdar, 2009).

$$D = [\prod_{i=1}^{n} c_i (y_i/x_i)]^{1/n} \tag{3}$$

Aggregation Metric D compares each organization's performance at two different states (i.e. current month compared to previous month). This comparison will provide an accurate report of the relative performance of the organization from one month to the next, but it doesn't enable a meaningful comparison between the two organizations, because even if the organizations are performing differently, the month-to-month comparison will only compare the organizations' previous month's performance. We illustrate a simple example in Table 1 that assumes a comparison between two similar organizations, where good performance is indicated by a higher percentage value. The illustration shows that despite an obvious difference in performance, the poor performing organization X actually reports a higher Aggregation Metric D value. Using the Aggregation metric D, as designed, we would rank the poor performing organization higher than the good performing organization, due to the comparison to system states relative only to each organization's respective previous month's performance.

Table 3-1. Aggregate Metric D illustration.

	Aggregate Metric D		
	Month 1	Month 2	Aggregate Value
Organization X Performance	55%	56%	102%
Organization Y Performance	100%	97%	97%

In the Table 1 illustration, we do not apply the c_i weight, because a linear weight in a multiplicative model doesn't influence the geometric distance between the metric values; it only serves to scale the product. This is another factor in our decision to pursue an alternative aggregation method, as we require the ability to differentiate between the importance and influence of individual metrics.

We chose to pursue using the geometric mean for aggregation, because simpler averaging methods like the arithmetic mean may not be able to meaningfully aggregate measurements in a system with the complexity of the nuclear enterprise (Kesheleva, 2009). Further, the geometric mean has advantages over more complex aggregation methods such as AHP. The geometric mean's main advantage over methods like AHP (in addition to simplicity) is that it is dimensionless and allows different units to be meaningfully aggregated (Sikdar, 2009). One of AHP's advantages is that it normalizes the data. The geometric mean also does this. Another advantage of the geometric mean is that it is always less than or equal to the arithmetic mean, which ensures that sensitivity to underperformance is selected for.

2.0 Aggregation *h* Method

We propose a unique method derived from the weighted geometric mean. The foundational assumptions for our research are as follows. We describe and demonstrate the Aggregation *h* method and use generate notional performance metric data, because the metrics do not currently exist and we wanted to test the sensitivity of the hierarchy and aggregation method by creating certain performance conditions for the metric data. We assumed that the metric data generated accurately represents real data. Also, we assumed that decision-makers prefer to review performance information in a condensed form versus viewing large numbers of metrics. We also assumed that the DoD definition of sustainment applies to nuclear enterprise sustainment.

Notation

h Aggregate value of input metrics to performance measurement hierarchy

h_i Value representing the normalized performance measure resulting from x_i and y_i comparisons

n Number of metrics $i = 1, \ldots, n$ describing a subcriterion

p Percent of metric representativeness in a subcriterion

w_i Weighting factor assigned to a given h_i

x_i Vector element measuring the actual performance of the i^{th} metric

y_i Vector element measuring the performance standard of the i^{th} metric

In pursuit of an aggregation method for nuclear enterprise metrics, we determined that a suitable aggregation method would require the capability to weight metrics, as well as compare performance between organizations and within the same organization over time. There are several techniques for weighting alternatives in multiple objective decision analysis; our method adapts the Value-focused Thinking additive value model method for weighting (Parnell, 2008). The second requirement, inter-organizational performance comparison, presented a challenge as we were unable to find a technique in the literature that met the specific requirements needed for aggregating logistics metrics.

The weighting system used in our model was adapted from the Value-focused Thinking additive value model, where the value of a given alternative is defined as the sum of the products of weights and alternatives, such that the weights for scoring a decision alternative sum to 1.0 (Parnell, 2008). However, since our model is multiplicative, we use a percent to represent the proportion each metric represents for a given tertiary subcriterion, where the percentages sum to 100 percent (or 1.0). The weight used in the aggregation calculation is the percent p_i for each metric times the number of metrics n in the tertiary subcriterion or n tertiary subcriteria in the subcriterion, shown in Figure 2. The process repeats when aggregating tertiary subcriteria into subcriteria, and so on. This method of weighting gives the decision-maker a simple task of assigning a percent to each metric, according to importance. We chose this method over Saaty's Analytic Hierarchy Process weighting method due to the simplicity. We set our metric, tertiary subcriteria and subcriteria weights where:

$$w_i = np_i \tag{4}$$

And

$$\sum_{i=1}^{n} p_i = 1 \tag{5}$$

Meaningful comparison of two or more similar organizations is a valuable tool for a decision-maker. Our aggregation method has this attribute. We determined that adding a performance standard for each metric, then comparing the metric to the standard—the metric is the numerator and the standard is the denominator—accomplished this goal. The resulting equation compares a vector of metrics x_i to a corresponding vector of performance standards y_i, which results in the ratio value h_i. The ratio value h_i is exponentially weighted w_i. The mean is determined by Equation 6. The weighting scheme is exponential, so the result has the effect of increasing the representativeness of the ratio h_i by w_i times, and since the root of the sum of w_i's is taken for the product, the mean is still representative of the constituent numbers.

The calculated h_i value for each metric is a normalized performance value, which allows it to be compared directly to any organization using the same metric. This is possible because the h_i value is no longer a metric value, but an absolute value of performance against a standard. Comparing it to another h_i value from a different organization will provide the decision-maker with a meaningful comparison of performance levels.

An important consideration in aggregating using this method is that the metrics must operate in the same direction. For example, for all metrics, an increase must indicate improvement or the converse. In our model an increase in a metric value indicates an improvement in performance.

We defined the ratio h_i to eliminate the possibility of ratio values greater than 1. This would occur when a metric x_i is greater than its standard y_i, and would cause two problems. First, having a range of aggregate values ranging from 0 to 1+ is difficult to interpret. It is customary to view performance measures where the ratios are bound to a range [0,1]. Second, the further the aggregate values are from one another, the less meaningful the aggregate value, particularly if the distance between metrics is in the upward direction. Simply put, if the aggregate value is allowed to exceed 1, the process will be less sensitive to downward movement, because the distance between the smallest and largest ratios will be greater (Kesheleva, 2009). For logistics performance, decision-makers are primarily concerned with performance up to a certain standard. Conversely, decision-makers are concerned when a subordinate organization is underperforming (i.e. their performance metrics do not meet the set standard).

Ideally, organizations should set the standard y_i at a value consistent with historical performance that meets organizational goals. We recommend that this value be established and subsequently adjusted using statistical process control techniques, such as p-charts or x-bar charts (Heizer, 2006).

$$\text{Aggregation } h = \left[\prod_{i=1}^{n} (h_i)^{w_i}\right]^{1/\sum_{i=1}^{n} w_i}, \text{ where } h_i = \begin{cases} x_i/y_i & \text{for } all\ x_i \le y_i \\ 1 & otherwise \end{cases} \quad (6)$$

Equation 6 is our final aggregation formula. The h_i calculation is performed on the metrics only. For tertiary subcriteria and subcriteria, h_i is set equal to the aggregate values being re-aggregated. We use the ratio comparison of metrics and standards only for metrics, because we are normalizing the data for performance comparison. The resulting values reflect absolute performance that we want to preserve in our aggregation up the hierarchy. The ratio comparison tightens up variance between good performing metrics and highlights the variance of poor performing metrics, which is preserved in the aggregation at each hierarchy level. This quality of the aggregation method is illustrated in our analysis.

2. Performance Measurement Hierarchy Construction

To construct a performance measurement system for nuclear sustainment, the strategic goal must be linked to outputs that can be directly measured. To determine strategically important outputs, a performance measurement hierarchy must be constructed.

2.1 Defining the Strategic Goal—Sustainment

The first step in creating a performance measurement hierarchy for nuclear enterprise sustainment was to carefully define the meaning of sustainment. We based the construction of the sustainment performance measurement hierarchy on the definition and description of sustainment found in Defense Acquisition Guidebook, Department of

Defense Instruction 5000.02, Operation of the Defense Acquisition System, and Department of Defense Directive 5000.01, The Defense Acquisition System. According to paragraph 3.9.2.1., the Defense Acquisition Guidebook defines sustainment as including supply, maintenance, transportation, sustaining engineering data management, configuration management, manpower/personnel, and training (Defense Acquisition Guidebook, 2010). The Defense Department directive expands this definition to include the life cycle from initial procurement to supply chain management (including maintenance), reutilization and disposal. It also emphasizes the importance of monitoring key support metrics (DoD directive 5000.01, 2003).

Department of Defense guidance, published in 2003, provides insight into how the sustainment phase of lifecycle management should be viewed. In particular, the guide links performance and sustainment, where performance is an indicator of sustainment operations and support investment (Eccles, 1991). In other words, system performance is a function of investment in lifecycle sustainment. Thus performance is the key measure of sustainment (Office of Secretary of Defense, 2003).

2.2 Performance Measurement Hierarchy Model

We used current academic literature (Analytic Hierarchy Process, value-focused thinking and the process method of performance measurement system design) to construct a strategy-linked performance measurement system for the sustainment of the nuclear enterprise. Also, in the interest of uniting our research with ongoing efforts by United States Air Force to measure performance of the nuclear enterprise, we incorporated feedback from more than a dozen United States Air Force nuclear enterprise leaders on hierarchy modeling.

We constructed the performance measurement hierarchy with the rationale that "performance measures need to be placed in a strategic context, as they influence what people do [and that] …measurement may be the process of quantification, but its effect is to stimulate action" (Neely, 1995).

Using sustainment as our strategic goal, the process method and feedback from nuclear enterprise leaders, we identify nine subcriteria that comprise the strategic goal: Weapons Storage Area Operations; Sustaining Engineering; Bomber Sustainment; Intercontinental Ballistic Missile (ICBM) Sustainment; Retirement and Disposal; Policy Performance; Support Equipment; Compliance; and Nuclear Infrastructure. To keep the scope of this paper manageable, we constructed the hierarchy for only Weapons Storage Area Operations. The other subcriteria would be developed the same way as the Weapons Storage Area Operation subcriterion. Also, as an aside, Weapons Storage Area Operations could be redefined to view enterprise performance of individual nuclear weapon systems (i.e. by bomb or warhead) versus enterprise performance of Weapons Storage Area Operations geographically (i.e. by base/unit).

Weapons Storage Area Operations is intended to measure the sustainment activities that take place in the Weapons Storage Area. Measuring the sustainment activities that take place in the Weapons Storage Area can act as a leading performance indicator to changes in capability. With meaningful Weapons Storage Area Operations measurements, leaders can make informed decisions on the allocation of scarce resources and act on negative trends to prevent serious incidents. Weapons Storage Area Operations should be thought of as analogous to elements of maintenance activities in United States Air Force backshop maintenance squadron and aircraft maintenance squadrons (Air Force Instruction 21-101, 2010). Although maintenance policy and technical guidance is different from other United States Air Force maintenance, the business processes are essentially the same.

Since the key business processes of Weapons Storage Area Operations can be seen as an analog to a United States Air Force aircraft maintenance, we use the comparison as a starting point to deviate from and to help communicate the Weapons Storage Area Operations subcriterion to United States Air Force leadership (Air Force Instruction 21-200, 2009; Air Force Instruction 21-101, 2010). This paper discusses specific metrics later; some of which are adapted from existing aircraft maintenance metrics, others are created to measure critical areas of Weapons Storage Area Operations not analogous to aircraft maintenance (U.S. Air Force Maintenance Metrics, 2009). It is important to emphasize that aircraft maintenance was *not* used as a template for this research, despite the adaptation of certain metrics, but primarily as a familiar reference point for consumers of this research.

We developed tertiary subcriteria for the Weapons Storage Area Operations subcriterion, based on feedback from nuclear enterprise leaders and personal experience. Weapons Storage Area Operations, as a subcriterion to sustainment, can be seen to have four tertiary subcriteria: Maintenance Performance, Stockpile Condition, Supply Chain Performance and Nuclear Expertise, as depicted in Figure 2.

The Maintenance Performance tertiary subcriterion is the aspect of Weapons Storage Area Operations most closely related to aircraft maintenance backshops. Maintenance Performance measures the performance of periodic maintenance activities conducted by United States Air Force personnel. The difference between nuclear maintenance and United States Air Force maintenance backshop maintenance lies mainly in policy and technical procedures, but the maintenance actions performed are the same as any organization performing periodic maintenance.

Stockpile Condition is the tertiary subcriterion that measures the condition of the nuclear stockpile in United States Air Force custody, as well as the key release gear associated to the weapons, as the condition of this equipment is considered essential to nuclear capability and is mated to weapons or warheads while in storage (Air Force Instruction 21-200, 2009).

Supply Chain Performance is comprised of both United States Air Force and Department of Energy supply activities. This tertiary subcriterion is intended to both capture the performance of the supply chain in sustaining the nuclear enterprise and to measure Nuclear Weapons Related Material policy compliance.

Finally, Nuclear Expertise is the fourth tertiary subcriterion. This subcriterion may seem out of place in the context of sustainment, but personnel are a part of the Department of Defense sustainment definition, as a technically competent workforce is essential to weapon system sustainment (Office of Secretary of Defense, 2003). Without trained and certified personnel, it is not possible to maintain the nuclear stockpile. People are a vital maintenance resource for field level nuclear sustainment and must be carefully managed and overseen to ensure a reliable nuclear stockpile (Air Force Instruction 21-200, 2009).

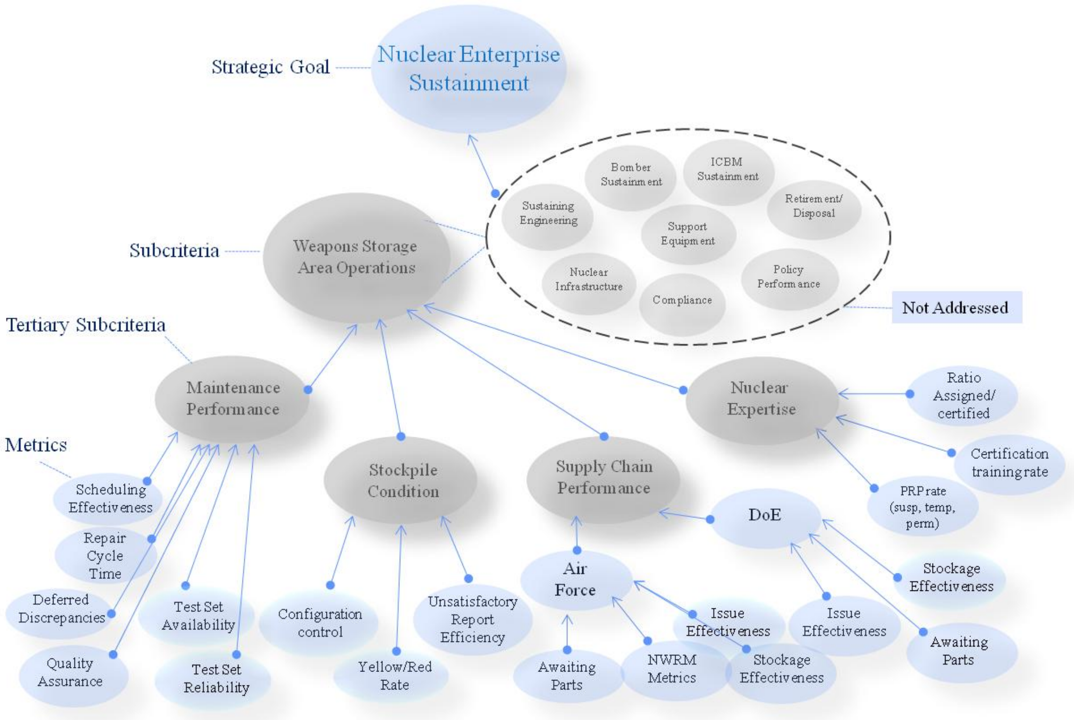


Figure 3-2. Nuclear sustainment performance measurement hierarchy.

We determined that the tertiary subcriteria-level of nuclear sustainment could be directly measured. The final step in hierarchy development was to identify the outputs or metrics that meaningfully describe the performance of the next higher level of the hierarchy, with traceability all the way up to the strategic goal, Sustainment. The metrics we describe, shown in Appendix A, attempt to measure the key business processes in Weapons Storage Area Operations (Air Force Instruction 21-200, Air Force Instruction 21-101 and Air Force Maintenance Metrics Handbook, 2009). We propose a minimum number of metrics that measure the timeliness and quality of the key business processes identified (Neely, 1995). The metrics identified for each tertiary subcriterion are organized in an index that allows meaningful aggregation (Silver, 2009). These metrics are not meant to be collectively exhaustive of all possible performance metrics, as there could be metrics legitimately added to more completely measure the subcriteria, but this should be approached cautiously, in a way that minimizes the total number of measurements (Neely, 1995).

4. Hierarchy Validation and Aggregation Sensitivity Analysis

There are two important considerations in performance measurement. First is that the organization must adequately define and communicated its strategic goals and that the resulting performance measurement hierarchy is meaningfully linked at every level. Success on this front will help ensure that the organization is measuring the right things and that the behavior of leaders at every hierarchical level is influenced to positively contribute to the strategic goal. Second, given a sound performance measurement hierarchy, it is of great importance the performance *information* is meaningfully conveyed to the decision-maker. In a complex, large organization, accurately communicating system performance is essential for the decision-maker to be able to make good decisions for the enterprise. We propose that using aggregation is a credible way to connect a quantitative "thread" from the raw metrics level through each level of the hierarchy. Our analysis shows that it is indeed possible to accurately capture system performance at every level of the hierarchy.

4.1 Hierarchy Validation

To analyze the sensitivity and benefit of the Aggregation *h*, we generated three sets notional metrics values, shown in raw for in Appendices B through D, intended to represent good, poor and mixed performance. We define good performance as metrics that are greater than or equal to 90 percent when compared to their corresponding standards. Poor performance is defined as metrics that are less than 80 percent when compared to their corresponding standards. Finally, for mixed performance, we set one metric in each tertiary subcriterion at a poor performing value that decreased between January and March, but then dramatically improves in April. The other metrics in each tertiary subcriterion for mixed performance were set to depict good performance, as defined.

First, we analyze the range and completeness of our sustainment hierarchy compared to the sustainment criteria recently developed by the United States Air Force A10 nuclear integration office, as shown in Table 3-2. Decomposing the detailed Department of Defense sustainment definition, we constructed a simple matrix to identify the areas our hierarchy measures and the areas the A10 office criteria measures.

Table 3-2. Unites States Air Force A10 office sustainment criteria.

A10 Office Sustainment Criteria
Provide available and serviceable Nuclear Certified Equipment
Maintain weapons storage areas and maintenance facilities
Maintain and track correct inventories of weapons, critical parts, and NWRM
Maintain responsive supply chain for bombers and ICBMs
Comply with NWRM handling/storage criteria
Perform sufficient number of weapon/weapon system operational tests
Perform adequate surveillance, assessment & certification and refurbishment of weapons

Table 3-3 shows a comparison of our hierarchy compared to the Unites States Air Force A10 office criteria. In our comparison, we intend to show the range and completeness of our model hierarchy compared to the A10 sustainment criteria. This comparison only addresses sustainment, which is only a part of the A10 performance measurement model. Our model addresses 21 of the 25 key elements of sustainment, whereas the A10 sustainment criteria address five elements.

Table 3-3. Sustainment hierarchy and A10 sustainment criteria comparison.

Sustainment Hierarchy Range		
Department of Defense sustainment elements	Sustainment Model Hierarchy	A10 Sustainment Criteria
Key support metrics	X	
Field Level Maintenance	X	
Depot Level Maintenance	X	
Disposal	X	
Retirment	X	
Sustaining Engineering	X	
Support Equipment	X	X
Supply	X	X
Inventory Management	X	X
Transportation		
Process Efficiency	X	
Supportability	X	
Reliability	X	
System Performance	X	
Maintainability	X	
Logistics IT		
Supply Chain Mangement	X	
Operations and Support	X	
Manpower and Personnel	X	X
Training	X	
Data Management		
Maintenance	X	
Environment and Habilitability		
Facilities	X	X
Maintenance Planning	X	

4.2 Aggregation Sensitivity Analysis

The following sensitivity analysis is a step-by-step illustration of the mechanics of our aggregation equation. The analysis starts with a detailed comparison of the metrics used to value the WSA Operations Subcriterion without our technique and with our technique. This comparison is followed by a demonstration of the aggregation process at each level of the nuclear sustainment hierarchy.

First, we show a side-by-side comparison of the raw metric values in our hierarchy against the same metrics after h_i is determined. Figures 3-3a and 3-3b through 3-5a and 3-5b illustrate the impact of the first step in our aggregation process. The Figures 3a, 3-4a and 3-5a show the variance between the raw metric values. However, when we calculate h_i, we significantly reduce the variance, as shown in Figures 3-3b, 3-4b and 3-5b. This reduction in variance allows us to more clearly see the true performance of the system, because the standards applied to the metric values allow us to compare an absolute measure of performance. This same quality allows comparison between different organizations, as long as the same metrics are used.

Figures 3-3a and 3-3b plot the Appendix B WSA Operations metric values and associated h_i ratio values for a scenario depicting good system performance. The dot markers show a visual illustration of the variance between the metrics for each tertiary subcriterion in the WSA Operations subcriterion. Figure 3-3a shows the raw metrics with values generated to depict good performance. The appearance of the spread between the markers shows significant variance between some of the individual metric values. Figure 3-3b shows the same metrics after h_i is calculated. This brings all the values into a tight cluster. Of note it allows a meaningful performance comparison between a reciprocal metric (metrics low on the vertical axis). This occurs because the h_i calculation compares the metric value to a standard, which results in a higher ratio value of performance.

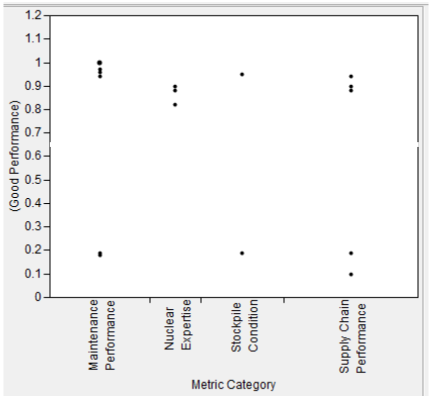


Figure 3-3a. Raw metrics for good performance.

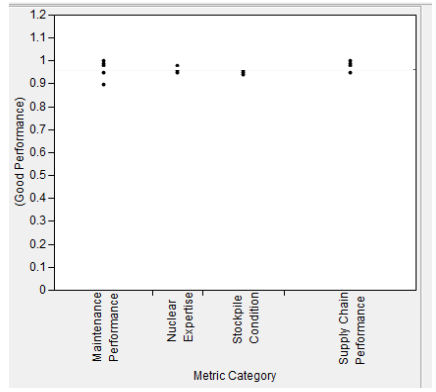


Figure 3-3b. *h* ratio metrics for good performance.

For a scenario depicting poor system performance, Figures 3-4a and 3-4b shows raw metric values, shown in Appendix C, for poor performance and associated h_i calculated ratio values. This illustration shows an image similar to the good performance example. This example indicates that the first step in the aggregation, calculating h_i, produces similar results. However, poor performance relative to the standard y_i necessarily results in h_i values less than one.

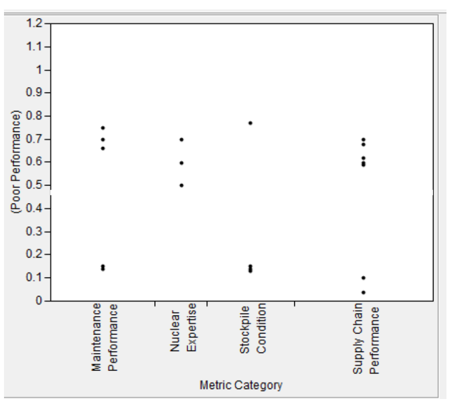


Figure 3-4a. Raw metrics for poor performance.

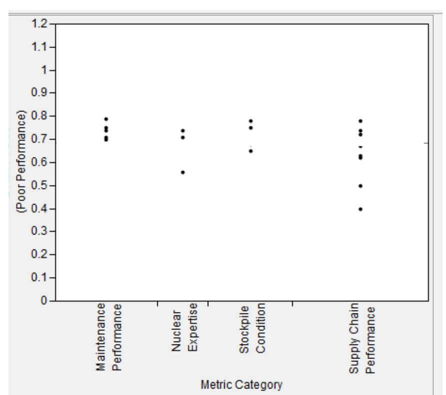


Figure 3-4b. *h* ratio metrics for poor performance.

Given that the h_i calculation produces similar results with consistent good or consistent poor performance, we decided to test the behavior using good performance with a single poor performing metric in each tertiary subcriterion to show what we are calling mixed performance. The results are shown in Figures 3-5a and 3-5b. The results are interesting, if predictable, in that we see significant variability in the raw metric values. However, after we calculate h_i, we observe in all four tertiary subcriteria shown in the same reduction of variance, but the poor performing metric becomes clearly evident, whereas it was not discernable in the raw metric chart. The outlying dots in the plot in Figure 3-5b can be referenced to the bold font metrics in Appendix D. Our assumption is that performance metric values in real-world scenarios would be of a mixed nature, where some show good performance and some show poor performance in a single category. The quality of h_i calculation to both tighten metric variance and highlight poor performance would be particularly useful.

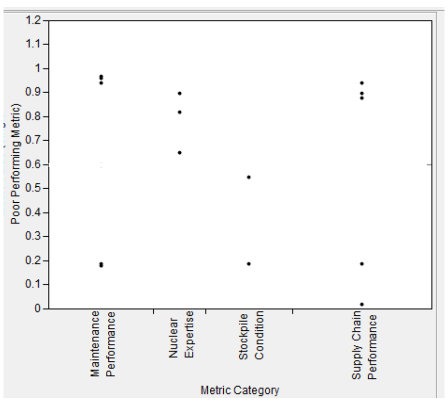


Figure 3-5a. One way analysis of raw metrics for mixed performance.

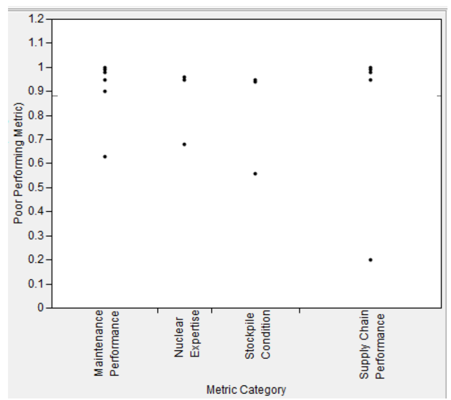


Figure 3-5b. *h* ratio metrics for mixed performance.

The next step in analysis and validation of the aggregation method is to illustrate the subsequent aggregation steps and explore the behavior of the metrics, tertiary subcriteria and subcriteria at each level of the hierarchy to determine if the aggregation meaningfully represents its constituents.

In Appendix B, WSA Operations metrics are shown for good performance (90 percent or higher for all metrics). The first column indicates the metric. The second and third columns show the percent weight of the metric (sums to 100 percent) for the tertiary subcriterion and the metric performance standard, respectively. The remaining columns show the raw monthly metric value then the associated h_i calculation for each metric.

Appendix C shows poor metric performance in all tertiary subcriteria in the WSA Operation subcriterion. However, we intentionally showed across the board improvement for the month of April to demonstrate the responsiveness or the aggregation method. The columns are organized the same way as the columns in Appendix B.

The raw metrics in Appendix D reflect mixed performance, marked by the steadily decreasing poor performance of a single metric followed by a dramatic improvement for the month of April. The metrics showing poor performance, bold font, are scheduling effectiveness, weapon yellow/red rate, USAF mission capable rate, and PRP certified rate.

In the first step of aggregation, Appendices A through C are used to perform organizational level aggregation WSA Operations' tertiary subcriteria. Careful comparison of the raw metrics to the aggregations shown in Tables 3-4 through 3-6 illustrates an accurate representation of performance at the organizational level aggregation. For our comparison, it is important to note that the three organizations can be characterized: good, poor and mixed (single poor performing metric).

Table 3-4. Organizational level aggregation for good performance.

Organization 1 (good performance)	Jan	Feb	Mar	Apr
Maintenance Performance	0.98	0.98	0.98	0.98
Stockpile Condition	0.95	0.94	0.95	0.97
Supply Chain Performance	0.97	0.99	0.99	0.98
Nuclear Expertise	0.97	0.99	0.97	0.98

Table 3-5. Organizational level aggregation for poor performance.

Organization 2 (poor performance)	Jan	Feb	Mar	Apr
Maintenance Performance	0.73	0.74	0.74	0.73
Stockpile Condition	0.72	0.70	0.68	0.66
Supply Chain Performance	0.60	0.59	0.62	0.68
Nuclear Expertise	0.63	0.63	0.60	0.59

Table 3-6. Organizational level aggregation for single poor performing metric.

Organization 3 (single poor performing metric)	Jan	Feb	Mar	Apr
Maintenance Performance	0.78	0.74	0.69	0.97
Stockpile Condition	0.82	0.80	0.81	0.97
Supply Chain Performance	0.63	0.61	0.58	0.98
Nuclear Expertise	0.81	0.78	0.79	0.98

Table 3-7 illustrates aggregation at the tertiary subcriteria level combining all three notional organizations: good, poor and mixed (single metric poor performance). The first columns indicate the tertiary subcriteria. The second column shows the percent weight for the tertiary for the next aggregation at the subcriteria level. The remaining columns display the aggregation of the three organizations' metrics in the indicated subcriteria. The aggregation reflects the mix of good and poor performance by showing a mid-point between the good and poor performing organizations, but the poor performing metric is also apparent, as it steadily decreases then shows marked improvement for the month of April.

Table 3-7. Tertiary subcriteria aggregation for three organizations.

WSA Operations Aggregation (three organizations)					
Tertiary Subcriteria	Percent	Jan	Feb	Mar	Apr
Maintenance Performance	25.00%	0.82	0.81	0.79	0.89
Stockpile Condition	25.00%	0.83	0.81	0.80	0.85
Supply Chain Performance	25.00%	0.72	0.71	0.71	0.87
Nuclear Expertise	25.00%	0.79	0.79	0.77	0.83

At this stage of the aggregation, it is appropriate to make a comparison to the two methods the Unites States Air Force now uses to present performance metrics to decision-makers. A common approach is simply to show the raw metrics, which would be equivalent to what we show in Appendices B through D, lacking a meaningful way to condense the data into decision-quality information. The other approach is to set triggers for metrics. This approach typically sets a performance floor for the metrics (red), perhaps some middling performance (yellow) and some reasonable range of good performance (green). These performance categories are triggered by the lowest performing metrics in a subcriterion (to use academic terminology).

Returning to our example using triggers, the following is a representation of what a United States Air Force decision-maker might be presented. We use the same data as shown in our aggregation example, up to this point. Presumably, all the metrics shown below would be red, simply because we take the reciprocal of a number of metrics where improvement is indicated by a decrease in value. This may appear to be an artificial problem introduced by our process. However, the alternative is to mix metrics that improve in different direction, which makes triggers an even more dubious method of measuring system performance. We selected the lowest performing metrics for each tertiary subcriterion. The key insight in this comparison is that seeing the poorest performance doesn't provide the decision-maker decision-quality information on the performance of the system at any level: organizational, tertiary subcriteria or even metric. In a complex organization, even though the decision-maker needs to be aware of weak areas, overall system performance is key because decision-makers need strategic information to allocate enterprise resources.

Table 3-8. Tertiary subcriteria displaying trigger metrics (poor performers).

Lowest performer trigger roll-up	Jan	Feb	Mar	Apr
Maintenance Performance	0.60	0.54	0.48	0.70
Stockpile Condition	0.55	0.54	0.54	0.71
Supply Chain Performance	0.02	0.02	0.02	0.10
Nuclear Expertise	0.50	0.52	0.51	0.50

For further consideration, recall that we intentionally generated metric values that emphasize obvious trends at the organizational level and we also placed values in the raw metrics for all three organizations a slight downward trend, ending in April with a sharp performance increase. Neither of these critical system performance insights is evident in Table 3-8. The consequence of making strategic decisions based on raw data (individual metrics) or a dangerously skewed roll-up, such as the one shown in Table 3-8, is misallocation of enterprise resources or target fixation on data points that don't reflect overall system performance (or where the system truly does need decision-maker focus). If we graphically compare our aggregation method against the lowest performing metric trigger method often used by the United States Air Force, it is evident that system performance is considerably different than the lowest performing raw metrics.

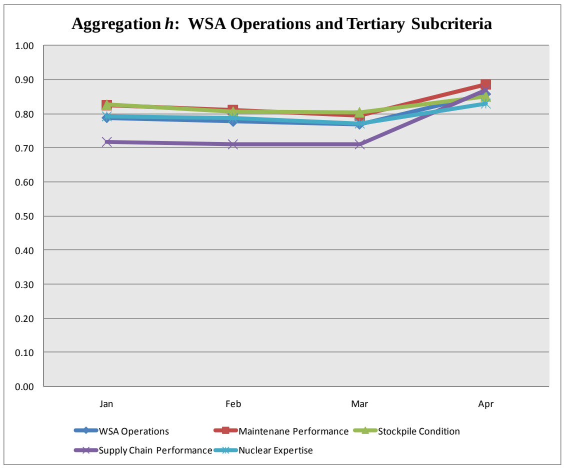


Figure 3-6. Aggregation of WSA Operations showing tertiary subcriteria.

When comparing Figures 3-6 and 3-7, Figure 3-6 indicates an overall higher level of performance—about 10 percent. Also, it is clear that reciprocal metrics (metrics where lower is better, we take the reciprocal to allow comparison) provide little insight into the tertiary subcriteria performance, let alone overall system performance. In the case of Nuclear Expertise, our system aggregation shows improvement in April, while the same data, as presented using the lowest performing metric, suggests a slight decrease.

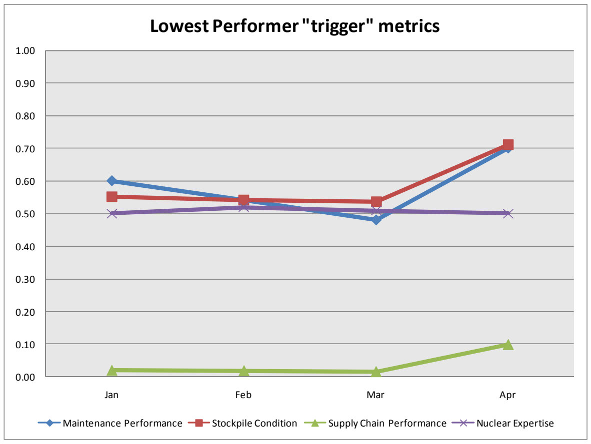


Figure 3-7. Lowest performing "trigger" metrics.

Table 3-9 shows aggregation at the subcriteria level. The subcriteria values not addressed in our research are arbitrarily set at 1. We weight the WSA Operations subcriteria for the strategic level aggregation in order to demonstrate the sensitivity of the method. Column one shows the subcriteria being aggregated. Column two is the percent weight for each subcriterion. The remaining columns show the aggregate value of each subcriterion, which itself is an aggregation of the tertiary subcriteria and the h_i value calculated from the raw metrics. Again, it is clear that the WSA Operations subcriterion aggregation reflects the constituent tertiary subcriterion values in Table 3-9. The steady decrease and marked increase of the poor performing metric organization can be detected at this level of aggregation.

Table 3-9. Subcriteria aggregation for three organizations.

Subcriteria	Percent	Jan	Feb	Mar	Apr
WSA Operations	60.00%	0.79	0.78	0.77	0.86
Nuclear Infrastructure	5.00%	1.00	1.00	1.00	1.00
Support Equipment	5.00%	1.00	1.00	1.00	1.00
Sustaining Engineering	5.00%	1.00	1.00	1.00	1.00
Policy Performance	5.00%	1.00	1.00	1.00	1.00
Retirement/ Disposal	5.00%	1.00	1.00	1.00	1.00
ICBM Sustainment	5.00%	1.00	1.00	1.00	1.00
Bomber Sustainment	5.00%	1.00	1.00	1.00	1.00
Compliance	5.00%	1.00	1.00	1.00	1.00

Finally, the strategic level aggregation for nuclear enterprise sustainment, Table 3-10, shows a less dramatic change than the tertiary subcriteria and subcriteria aggregations, but the behavior of the constituents of the aggregation is still apparent.

Table 3-10. Strategic level goal aggregation for nuclear enterprise sustainment.

Strategic goal	Jan	Feb	Mar	Apr
Nuclear Enterprise Sustainment	0.87	0.86	0.85	0.91

5. Conclusions

Performance measurement theory emphasizes the importance of creating a performance measurement system that links strategic goals with the metrics the organization uses to measure success. If the strategic goal and metrics are aligned, it is likely that managers at all levels will be influenced to positively contribute to the organization's strategic goals. Additionally, in large complex organizations, it is important to be able to turn metrics data into information that decision-makers can readily understand and act upon.

By applying the aggregation method demonstrated in this paper, it is possible to provide a decision-maker with an accurate picture of organizational health at every level and for every critical business process. The alternatives to meaningfully aggregating performance metrics is to present a decision-maker with raw metrics data or establish trigger points that highlight poor performance. These alternatives plague the decision-maker with the burden of sifting through a sea of metrics or relying on a single data point to make informed decisions for the organization. We demonstrate a method of aggregation that can effectively provide insight into holistic view of performance that may contribute to more efficient and better strategic decision-making.

Using the process approach to performance measurement hierarchy construction and using the Department of Defense definition of sustainment, we found consistent feedback between leaders in the nuclear enterprise with respect to the subcriteria, tertiary subcriteria and metrics that should used to measure the performance of nuclear enterprise sustainment. We conclude that starting with a strategic goal that is both clearly defined and has institutional meaning was the basis for the consistent agreement among leaders at many levels. Further, we assert that differences between our hierarchy and the measurement efforts by various United States Air Force staff offices is rooted in our theoretical approach: a carefully defined sustainment goal and the deliberate linkage of the strategic goal to each level of the hierarchy.

6. Recommendations

The final form of the sustainment performance measurement hierarchy should be considered a foundation, or starting point, for senior decision-makers to use for operationalization of a nuclear enterprise performance measurement system. At the metric level, we adopted or adapted accepted United States Air Force metrics for measuring key business processes. This level of the hierarchy is somewhat subjective, though there was no dissent from leaders interviewed. We believe changes to the metrics level of the hierarchy would likely be to *add* metrics and there may, indeed, be a valid cause to do so. However, we submit one final caution concerning metrics, and performance measurement, generally. If we use too many or the wrong metrics, we diminish the ability of the decision-maker to accurately assess organizational health, we sub-optimize organizational performance and obscure the path toward the strategic goal.

Finally, we found that using our Aggregation *h* method can meaningfully communicate organizational performance at multiple levels in a performance measurement hierarchy. The benefits stem from its simplicity and the quality it has of being able to compare different organizations, given the same business process measurements.

We recommend further research to analyze the effectiveness of the metrics we developed and validation of the key business processes we identified to measure weapons storage area operations. Additionally, significant research is required to develop tertiary subcriteria and metrics for the eight other subcriteria not addressed in our sustainment hierarchy. With respect to our Aggregation *h* method, we recommend applying equation to other organizational performance measurement hierarchies. Also, we believe that the method could be further enhanced by setting variance thresholds at each level of aggregation to allow decision-makers to accurately and quantitatively determine which metrics, tertiary subcriteria and subcriteria are influencing organizational performance. In this way, decision-makers could identify the most beneficial areas to apply scarce resources.

Appendix A

Weapons Storage Area Operations Metrics

Maintenance Performance		
Scheduling Effectiveness	(number of completed events)/(total events scheduled) X 100	The primary aim of sustainment at the unit level is periodic maintenance management. Accomplishing periodic maintenance on-time and as scheduled is an important indicator of management's ability to plan resource allocation. Scheduling effectiveness also provides insight into the health of the unit's training and certification program, because accomplishing scheduled work relies on limited variability of repair cycle time and certified team efficiency
Repair Cycle Time	(Total hours per weapon, system, package)/(number of weapons, systems, packages)	Repair cycle time is a common measure in most production activities. Repair cycle time provides insight into process efficiency, as well as the skill and adequacy of the labor force. For nuclear sustainment, repair cycle time also indirectly indicates the quality of technical and engineering support.
Deferred Discrepancies	Total deferred events/total assigned weapons (includes all deferred events on weapons, release gear, handling equipment)	Tracking deferred maintenance goes hand-in-hand with scheduling effectiveness. As with aircraft maintenance, managing the number of deferred maintenance events is important to the health of the stockpile. Additionally, tracking deferred maintenance ensures a check and balance is in place for maintenance scheduling.
Quality Assurance	(Number of Quality Verification Inspections passed)/(Total Quality Verification Inspections) X 100	The Quality Assurance metric measures the quality of business processes ranging from nuclear warhead maintenance and technical guidance adherence to maintenance data collection accuracy and supply management. This measure coupled with measures like Repair Cycle Time, Scheduling Effectiveness and Deferred Discrepancy rate show the management's ability to efficiently use human and material resources while maintaining the highest possible maintenance management standards.
Test Set Availability	(Total operational hours)/(total hours) X 100	Nuclear enterprise sustainment relies heavily on nuclear certified test set reliability. Measuring test set availability, combined with other measures provides insight into repair cycle time, yellow/red rate, scheduling effectiveness and deferred maintenance.
Test Set Reliability	(Total number of test fails)/(total number of test events) X 100	Along with test set availability, test set failures are important to measure, because failures result in a significant contribution to repair cycle time and scheduling effectiveness. Also, test set availability does not capture many failures that impact maintenance efficiency, because test set operational hours aren't impacted by test failures.
Stockpile Condition		
Configuration Control: Time Compliance Technical Order (TCTO) and Retrofit Order (RO) Compliance	(TCTO/RO completed)/(TCTO/RO required) X 100	This metric measures configuration control, primarily measured by compliance with TCTOs/Ros, for nuclear weapons and key equipment. Configuration control is an important element of stockpile reliability.
Unsatisfactory Report (UR) Turn-Time	# of URs over 30 days/total URs	The UR process is a technical review process that requires inter-organization coordination and communication. Measuring UR turn time is important, because URs can impact the flow of periodic maintenance.
Yellow/Red Rate	(total red weapons)/(total accountable weapons) X 100	The yellow/red rate is a lagging performance measurement, much like mission capable is for aircraft mx. It provides insight to overall stockpile health, as well as mx efficiency and the quality of technical and engineering support. This rate should be relatively low. If it is less than 100%, other metrics might provide insight into this downward movement in this metric. For example, UR turn time may be a leading indicator to this weapons capability rate.

Supply Chain Performance (USAF and DoE)		
Nuclear Issue Effectiveness	(issues)/(issues and backorders) X 100	Issue effectiveness is a measure of how well logistics is supporting the customer. It measures any request to supply, not just requests for authorized items (items stocked). It is usually lower than stockage effectiveness, but is considered more representative of the customer's point of view.
Nuclear Stockage Effectiveness	(issues)/(issues and backorders minus unauthorized backorders) X 100	Stockage effectiveness measures the percentage of customer request filled by supply for items authorized to stock. Since supply can't stock every part, only the most frequently requisition or critical parts are authorized to stock. This metric measures supply and depot capability to manage demand for these items.
Issue Effectiveness	(issues)/(issues and backorders) X 100	Issue effectiveness is a measure of how well logistics is supporting the customer. It measures any request to supply, not just requests for authorized items (items stocked). It is usually lower than stockage effectiveness, but is considered more representative of the customer's point of view.
Stockage Effectiveness	(issues)/(issues and backorders minus unauthorized backorders) X 100	Stockage effectiveness measures the percentage of customer request filled by supply for items authorized to stock. Since supply can't stock every part, only the most frequently requisition or critical parts are authorized to stock. This metric measures supply and depot capability to manage demand for these items.
Awaiting Parts (AWP)	(# of AWP)/(total weapons stockpile)	AWP is the average number of parts backordered across the stockpile.
Nuclear Weapons Related Material (NWRM) Metrics	As published in Nuclear Logistics Surety document.	There are a number of existing NWRM metrics that measure the United States Air Force's ability to control and maintain visibility of NWRM items in the supply system.

Nuclear Expertise		
Certified Technicians	(# certified on tasks)/(# of assigned personnel) X 100	This metric captures a critical element of nuclear sustainment at the field level. Certified technicians are essential to performing periodic maintenance and maintaining a reliable stockpile. The maintenance capability letter (MCL) is the list of tasks for which a unit is required to maintain certified personnel. The ratio of certified to assigned personnel is a good gauge of the utilization of human resources, the effectiveness of the unit's training program and it's ability to efficiently perform required maintenance.
Certification Training Rate	(# days training for cert)/(# days scheduled for cert training) X 100	Certification training throughput is an important measure of a unit's training quality and management oversight of human resources. Certification training can take up to a year for a newly assigned Airman. It is important to control variance in the training schedule to ensure continuity of the training process and to ensure competent technicians are available to perform nuclear maintenance. If variance exists in the training process, or if units have significantly different throughout rates, management should determine the reason. Certification shouldn't be rushed, but it must also be managed aggressively and requires a project management approach to ensure a viable program.
Personnel Reliability Program (PRP) Certification Rate	(# of suspended, temporary decertified, permanent decertified)/(# of personnel on PRP) X 100	Like nuclear maintenance task certification, PRP certification is an essential part of nuclear maintenance. PRP certification rates should be monitored to ensure the number of suspended, temp and permanently decertified doesn't start to impact the flow of maintenance. Personnel suspended or decertified from PRP are not available to perform nuclear maintenance. In fact, they can consume more resources, because they must be escorted. The net effect of suspension and decertification is a reduction in maintenance capability. The purpose of the PRP program is to ensure high reliability of the people who work on or have access to nuclear weapons, and the commander must work to ensure squadron personnel and support organizations understand the program. For example, even administrative inefficiency can result in unnecessary time suspended for personnel who seek routine medical care. If interagency communication is not efficient, a suspended person may remain so only because of administrative inefficiency.

Appendix B

Raw Metrics for Good Performance

Maintenance Performance Tertiary Subcriterion	Percent	Standard	Jan	*h* ratio value	Feb	*h* ratio value	Mar	*h* ratio value	Apr	*h* ratio value
Scheduling Effectiveness	50.00%	0.95	1.00	1.00	0.98	1.00	0.97	1.00	0.99	1.00
Repair Cycle Time	10.00%	0.20	0.18	0.90	0.19	0.95	0.19	0.95	0.19	0.95
Deferred Discrepancies	10.00%	0.20	0.19	0.95	0.19	0.95	0.19	0.95	0.18	0.90
Quality Assurance	10.00%	0.95	0.96	1.00	0.94	0.99	0.93	0.98	0.93	0.98
Test Set Availability	10.00%	0.95	0.94	0.99	0.92	0.97	0.91	0.96	0.92	0.97
Test Set Reliablity	10.00%	0.99	0.97	0.98	0.96	0.97	0.97	0.98	0.99	1.00
Stockpile Condition Tertiary Subcriterion	Percent	Standard	Jan	*h* ratio value	Feb	*h* ratio value	Mar	*h* ratio value	Apr	*h* ratio value
Configuration Control--TCTO	18.00%	0.20	0.19	0.95	0.18	0.90	0.19	0.93	0.19	0.95
Configuration Control--RO	18.00%	0.20	0.19	0.95	0.18	0.90	0.19	0.93	0.19	0.94
Weapon Yellow/Red Rate	27.00%	0.99	0.95	0.96	0.98	0.99	0.97	0.98	0.97	0.98
UR Turn-Time	19.00%	0.20	0.19	0.95	0.19	0.95	0.19	0.94	0.20	0.98
107 Request Turn-time (ETAR)	18.00%	0.20	0.19	0.94	0.19	0.94	0.19	0.95	0.20	0.98
Supply Chain Performance Tertiary Subcriterion	Percent	Standard	Jan	*h* ratio value	Feb	*h* ratio value	Mar	*h* ratio value	Apr	*h* ratio value
USAF Awaiting Parts	9.00%	0.90	0.90	1.00	0.89	0.99	0.89	0.99	0.89	0.99
USAF Stockage Effectiveness	9.00%	0.90	0.88	0.98	0.89	0.99	0.89	0.99	0.90	1.00
USAF Issue Effectiveness	9.00%	0.95	0.90	0.95	0.94	0.99	0.93	0.98	0.94	0.99
USAF MICAP RATE	27.00%	0.10	0.10	0.99	0.10	1.00	0.10	1.00	0.10	0.99
USAF NWRM	18.00%	0.99	0.94	0.95	0.95	0.96	0.96	0.97	0.97	0.98
DoE Stockage Effectiveness	10.00%	0.95	0.94	0.99	0.96	1.00	0.95	1.00	0.92	0.97
DoE Issue Effectiveness	9.00%	0.95	0.90	0.95	0.94	0.99	0.93	0.98	0.91	0.96
DoE Awaiting Parts	9.00%	0.20	0.19	0.95	0.20	0.98	0.19	0.96	0.20	1.00
Nuclear Expertise Tertiary Subcriterion	Percent	Standard	Jan	*h* ratio value	Feb	*h* ratio value	Mar	*h* ratio value	Apr	*h* ratio value
Certified/Assigned Technicians	25.00%	0.85	0.82	0.96	0.85	1.00	0.84	0.99	0.84	0.99
PRP Certified Rate	50.00%	0.90	0.88	0.98	0.88	0.98	0.85	0.94	0.88	0.98
Task Certification Throughput Rate	25.00%	0.95	0.90	0.95	0.96	1.00	0.94	0.99	0.94	0.99

Appendix C

Raw Metrics for Poor Performance

Maint. Performance Tertiary Subcriterion	Percent	Standard	Jan	h ratio value	Feb	h ratio value	Mar	h ratio value	Apr	h ratio value
Scheduling Effectiveness	50.00%	0.95	0.70	0.74	0.72	0.76	0.72	0.76	0.71	0.75
Repair Cycle Time	10.00%	0.20	0.14	0.70	0.14	0.70	0.14	0.70	0.15	0.75
Deferred Discrepancies	10.00%	0.20	0.15	0.75	0.14	0.70	0.14	0.70	0.14	0.70
Quality Assurance	10.00%	0.95	0.66	0.69	0.65	0.68	0.66	0.69	0.64	0.67
Test Set Availability	10.00%	0.95	0.75	0.79	0.74	0.78	0.71	0.75	0.68	0.72
Test Set Reliablity	10.00%	0.99	0.70	0.71	0.70	0.71	0.72	0.73	0.70	0.71
Stockpile Condition Tertiary Subcriterion	Percent	Standard	Jan	h ratio value	Feb	h ratio value	Mar	h ratio value	Apr	h ratio value
Configuration Control--TCTO	18.00%	0.20	0.13	0.65	0.13	0.65	0.14	0.70	0.14	0.68
Configuration Control--RO	18.00%	0.20	0.14	0.68	0.13	0.65	0.13	0.63	0.13	0.64
Weapon Yellow/Red Rate	27.00%	0.99	0.77	0.78	0.75	0.76	0.72	0.73	0.71	0.72
UR Turn-Time	19.00%	0.20	0.15	0.75	0.14	0.70	0.13	0.65	0.12	0.60
107 Request Turn-time (ETAR)	18.00%	0.20	0.15	0.75	0.14	0.70	0.14	0.70	0.13	0.65
Supply Chain Performance Tertiary Subcriterion	Percent	Standard	Jan	h ratio value	Feb	h ratio value	Mar	h ratio value	Apr	h ratio value
USAF Awaiting Parts	9.00%	0.90	0.60	0.67	0.60	0.67	0.61	0.68	0.70	0.78
USAF Stockage Effectiveness	9.00%	0.90	0.70	0.78	0.70	0.78	0.68	0.76	0.70	0.78
USAF Issue Effectiveness	9.00%	0.95	0.59	0.62	0.55	0.58	0.60	0.63	0.62	0.65
USAF MICAP RATE	27.00%	0.10	0.05	0.50	0.05	0.50	0.06	0.60	0.07	0.70
USAF NWRM	18.00%	0.99	0.62	0.63	0.60	0.61	0.55	0.56	0.58	0.59
DoE Stockage Effectiveness	10.00%	0.95	0.62	0.65	0.62	0.65	0.66	0.69	0.68	0.72
DoE Issue Effectiveness	9.00%	0.95	0.68	0.72	0.62	0.65	0.64	0.67	0.67	0.71
DoE Awaiting Parts	9.00%	0.20	0.10	0.50	0.10	0.50	0.11	0.55	0.12	0.60
Nuclear Expertise Tertiary Subcriterion	Percent	Standard	Jan	h ratio value	Feb	h ratio value	Mar	h ratio value	Apr	h ratio value
Certified/Assigned Technicians	25.00%	0.85	0.60	0.71	0.58	0.68	0.54	0.64	0.55	0.65
PRP Certified Rate	50.00%	0.90	0.50	0.56	0.52	0.58	0.51	0.57	0.50	0.56
Task Certification Throughput Rate	25.00%	0.95	0.70	0.74	0.65	0.68	0.60	0.63	0.58	0.61

Appendix D

Raw Metrics for Mixed Performance

Maintenance Performance Tertiary Subcriterion	Percent	Standard	Jan	*h* ratio value	Feb	*h* ratio value	Mar	*h* ratio value	Apr	*h* ratio value
Scheduling Effectiveness	50.00%	0.95	0.60	0.63	0.54	0.57	0.48	0.51	0.93	0.98
Repair Cycle Time	10.00%	0.20	0.18	0.90	0.18	0.90	0.18	0.90	0.19	0.95
Deferred Discrepancies	10.00%	0.20	0.19	0.95	0.19	0.95	0.19	0.95	0.19	0.95
Quality Assurance	10.00%	0.95	0.96	1.00	0.94	0.99	0.93	0.98	0.93	0.98
Test Set Availability	10.00%	0.95	0.94	0.99	0.92	0.97	0.91	0.96	0.92	0.97
Test Set Reliablity	10.00%	0.99	0.97	0.98	0.96	0.97	0.96	0.97	0.99	1.00
Stockpile Condition Tertiary Subcriterion	Percent	Standard	Jan	*h* ratio value	Feb	*h* ratio value	Mar	*h* ratio value	Apr	*h* ratio value
Configuration Control--TCTO	18.00%	0.20	0.19	0.95	0.18	0.90	0.19	0.93	0.19	0.95
Configuration Control--RO	18.00%	0.20	0.19	0.95	0.18	0.90	0.19	0.93	0.19	0.94
Weapon Yellow/Red Rate	27.00%	0.99	0.55	0.56	0.54	0.55	0.53	0.54	0.97	0.98
UR Turn-Time	19.00%	0.20	0.19	0.95	0.19	0.95	0.19	0.94	0.20	0.98
107 Request Turn-time (ETAR)	18.00%	0.20	0.19	0.94	0.19	0.94	0.19	0.95	0.20	0.98
Supply Chain Performance Tertiary Subcriterion	Percent	Standard	Jan	*h* ratio value	Feb	*h* ratio value	Mar	*h* ratio value	Apr	*h* ratio value
USAF Awaiting Parts	9.00%	0.90	0.90	1.00	0.84	0.93	0.83	0.92	0.89	0.99
USAF Stockage Effectiveness	9.00%	0.90	0.88	0.98	0.88	0.98	0.85	0.94	0.90	1.00
USAF Issue Effectiveness	9.00%	0.95	0.90	0.95	0.87	0.92	0.87	0.92	0.94	0.99
USAF MICAP RATE	27.00%	0.10	0.02	0.20	0.02	0.19	0.02	0.16	0.10	0.98
USAF NWRM	18.00%	0.99	0.94	0.95	0.93	0.94	0.92	0.93	0.97	0.98
DoE Stockage Effectiveness	9.00%	0.95	0.94	0.99	0.93	0.98	0.92	0.97	0.92	0.97
DoE Issue Effectiveness	9.00%	0.95	0.90	0.95	0.90	0.95	0.90	0.95	0.93	0.98
DoE Awaiting Parts	10.00%	0.20	0.19	0.95	0.19	0.95	0.19	0.95	0.20	1.00
Nuclear Expertise Tertiary Subcriterion	Percent	Standard	Jan	*h* ratio value	Feb	*h* ratio value	Mar	*h* ratio value	Apr	*h* ratio value
Certified/Assigned Technicians	25.00%	0.85	0.82	0.96	0.80	0.94	0.80	0.94	0.84	0.99
PRP Certified Rate	50.00%	0.95	0.65	0.68	0.64	0.65	0.63	0.66	0.93	0.98
Task Certification Throughput Rate	25.00%	0.95	0.90	0.95	0.90	0.95	0.90	0.95	0.94	0.99

Appendix E
Blue Dart

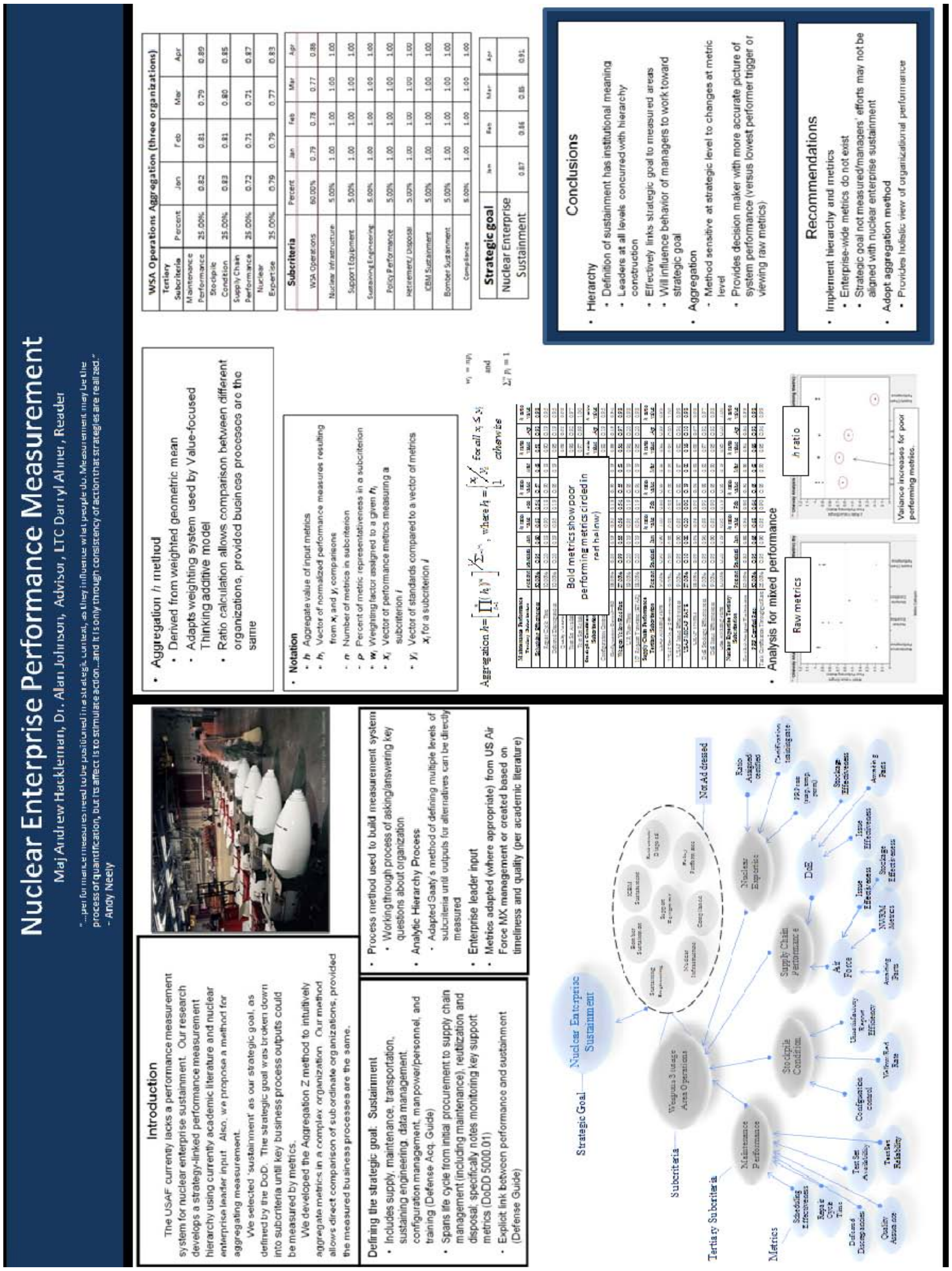
The criticality of the United States Air Force nuclear enterprise demands that commanders have the best possible understanding of system performance, both in the aggregate and at the drill-down levels sufficient to make timely corrective actions when warranted. We model a strategy-linked measurement system for nuclear enterprise sustainment. We propose a new Aggregation *h* method for aggregating performance metrics using United States Air Force approved or adapted metrics that possess the capability to weight metrics, as well as compare performance between organizations and within the same organization over time. We demonstrate our method with generated performance data designed to test the sensitivity of our method. Our Aggregation *h* method provides a simple, intuitive measurement approach that enables unity of effort and influences behavior at each hierarchical level towards achieving strategic goals, and is extendable to performance measurement for other complex sustainment systems.

Our results provide a solid foundation for performance measurement of nuclear enterprise sustainment. Using the Department of Defense definition of sustainment and mapping the key definitional elements to key business process outputs, we produce a strategy-linked performance measurement hierarchy, which provides the nuclear enterprise with a framework to use as a starting point for enterprise performance measurement.

In addition to constructing a performance measurement hierarchy, we demonstrated the efficacy of performance metric aggregation using our Aggregation *h* method. We show that aggregation at hierarchical levels can provide decision-makers with accurate system performance information currently lacking in Air Force performance measurement systems. Accurate information on system performance can enable decision-makers to make the best possible decisions with respect to the allocation of enterprise resources.

Appendix F

Quad Chart



Bibliography

Air Force Instruction 10-201. (2006). *Status of resources and training system*

Air Force Instruction 21-101. (2010). *Aircraft and equipment maintenance management*

Air Force Instruction 21-200. (2009). *Munitions and missile maintenance management*

Air Force Logistics Management Agency. (2009). *U.S. air force maintenance metrics.* Air Force Journal of Logistics.

Brignall, S., & Modell, S. (2000). An institutional perspective on performance measurement and management in the 'new public sector'. *Management Accounting Research, 11*, 281-306. doi:10.1006/mare.2000.0136

Chan, F. T. S. (2003). Performance measurement in a supply chain. *International Journal of Advanced Manufacturing Technology*, (21), 534-548.

Defense Acquisition University. (2010). *Defense acquisition guidebook.* Retrieved November 15, 2010, from https://acc.dau.mil/CommunityBrowser.aspx?id =314767&lang=en-US

Department of Defense Directive 5000.01: The Defense Acquisition System, DODDU.S.C. (2003).

Eccles, R. G. (1991). The performance measurement manifesto. *Harvard Business Review, January-February*, 131-137.

Gunasekaran, A., Patel, C., & McGaughey, R. E. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics*, (87), 333-347.

Headquarters United States Air Force. (2009). *Nuclear logistics surety implementation plan.* Unpublished manuscript.

Headquarters United States Air Force. (2008). *Blue Ribbon Review: Nuclear Weapons Policies and Procedures. Executive Summary.* Unpublished manuscript.

Johnson, J. P. (2007). *Balanced Scorecard: Aggregating Aircraft Mission Capabilities.* Unpublished Masters Thesis, Air Force Institute of Technology, DTIC.

Kosheleva, O., & Kreinovich, V. (2009). Guesstimation: A new justification of the geometric mean heuristic. *Applied Mathematical Sciences, 3*(47), 2335-2342-8.

Maj Gen Close, Kathleen D., & Lumpkins, S. A. (2010). *Nuclear logistics transition plan.*

Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: A literature review and research agenda. *International Journal of Operations and Production Mangement, 15*(4), 80-116.

Office of Secretary of Defense. (2003). *Designing and assessing supportability in DOD weapon systems: A guide to increased reliability and reduced logistics footprint* .

Office of Secretary of Defense. (2008). "Task Force on DOD Nuclear Weapons Management."

Office of the Under Secretary of Defense for Acquisition, Technology and Logistics. *The defense science board permanent task force on nuclear weapons surety: Report on the unauthorized movement of nuclear weapons.* Washington, D.C., 20301-3140: Defense Science Board, Department of Defense. Retrieved from http://www.acq.osd.mil/dsb/reports/ ADA480063.pdf

Parnell, Gregory S. (2008). Multiobjective Decision Analysis. John Wiley & Sons, Inc. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/ 9780470087923. hhs010/full

Pendley, S. A., Thoele, B. A., Albrecht, T. W., Howe, J. A., Antoline, A. F., & Golden, R. D. (2007). Aligning maintenance metrics: Improving C-5 TNMCM. *Air Force Journal of Logistics, 37*(1), 22 October 2010-12-22. Retrieved from http://www.aflma.hq.af.mil/shared/media/document/AFD-100111-023.pdf

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operations Research*, *48*(1), 9-26.

Scott A. Haines, C. (2009). Capabilities-based resourcing for air force weapon system sustainment: Efficiency versus effectiveness.

Sikdar, S. K. (2009). On aggregating multiple indicators into a single metric for sustainability. *Clean Technologies and Environmental Policy, 4*(3), 16 October 2010. doi:10.1007/s10098-009-0225-4

Silver, M. (2009). *An index number formula problem: The aggregation of broadly comparable items.* International Monetary Fund.

The Centre for Business Performance, Cranfield School of Management. (2004). *Literature review on performance measurement and management*The IDeA and Audit Commission Performance Management, Measurement and Information (PMMI) Project.

The White House. (May 2010). *National security strategy*

Vita

Major Andrew S. Hackleman is an aircraft and munitions maintenance officer. He enlisted in the Air Force in December 1996 and commissioned in January 2001. Major Hackleman's assignments in maintenance squadrons at Minot Air Force Base, Nellis Air Force Base, Creech Air Force Base, Dyess Air Force Base, Whiteman Air Force Base and Wright-Patterson Air Force Base. Follow graduation from the Air Force Institute of Technology, he is being assigned to the Nuclear Weapons Center at Kirtland Air Force Base.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
03/24/2011	**Master's Thesis**	Sep 2009 - Mar 2011

4. TITLE AND SUBTITLE
NUCLEAR ENTERPRISE PERFORMANCE MEASUREMENT

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hackleman, Andrew, Major, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Street, Building 642
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-LSCM-ENS-11-05

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Mr. Gregory Gross and Lt Col Kenneth Bottari
AIR FORCE NUCLEAR WEAPONS CENTER
1551 Wyoming Blvd SE
Kirtland AFB NM 87117-5624

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The criticality of the United States Air Force nuclear enterprise demands that commanders have the best possible understanding of system performance, both in the aggregate and at the drill-down levels sufficient to make timely corrective actions when warranted. We model a strategy-linked measurement system for nuclear enterprise sustainment. We propose a new Aggregation *h* method for aggregating performance metrics using United States Air Force approved or adapted metrics that possess the capability to weight metrics, as well as compare performance between organizations and within the same organization over time. We demonstrate our method with generated performance data designed to test the sensitivity of our method. Our Aggregation *h* method provides a simple, intuitive measurement approach that enables unity of effort and influences behavior at each hierarchical level towards achieving strategic goals, and is extendable to performance measurement for other complex sustainment systems.

15. SUBJECT TERMS
Performance measurement, Aggregation, Metrics

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Alan W. Johnson, PhD, ENS
U	**U**	**U**	**UU**	83	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 4703 (Alan.Johnson@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18