

DIAGNOSING PERFORMANCE MANAGEMENT AND PERFORMANCE BUDGETING SYSTEMS: A CASE STUDY OF THE U.S. NAVY

Natalie J. Webb
Defense Management Resources Institute,
Monterey, California

Philip J. Candreva
Naval Postgraduate School,
Monterey, California

ABSTRACT

We present here a case study of an organization within the U.S. Navy that created a new organizational construct and performance management system. We explore the issues faced by naval leaders as they attempt to use their performance information to make resource allocation decisions at the sub-organization level, and drive budgets at the organization and service (navy) level. We diagnose the practical problems a government organization encounters when implementing a performance management system, to include their influence on budgets, and make recommendations for public sector performance budgeting organizations. This case confirms challenges noted in the literature associated with performance management and performance budgeting systems. We offer recommendations for public officials considering such endeavors.

1. INTRODUCTION

In a recent manuscript, Schick (2008, p. 2) states, “the literature and practices of performance budgeting have been too long on exhortation and too short on diagnosis”. We present here a diagnostic case of a performance management system in the U.S. Navy, examined through the perspective of those exhortations. We assess the strengths and weaknesses of the system

Public Finance and Management
Volume 10, Number 3, pp. 524-555
2010

and the issues faced by naval leaders as they attempt to use their performance information to formulate budget requests and execute budgets. We diagnose many of the practical problems a government organization encounters in designing and using a performance management system, especially when it seeks to extend the system to performance budgeting. In so doing, we provide empirical evidence of many of the findings in the literature, contribute to the understanding of how a performance-based management system directs managers, and how such a system can (or cannot) be used to implement a performance-based budgeting system.

We examine the surface warfare enterprise (SWE), an organizational construct that is part of a larger “Navy Enterprise” initiative.¹ Broadly, the SWE is a construct that seeks to link various organizations involved in policy decisions and implementation of policy, including defining needs for and constructing, operating, and employing naval surface ships. We concentrate on the Naval Surface Force and its role in manning, training, equipping, and sustaining the existing surface fleet of 162 ships. The ultimate outcome for navy ships is how they perform if and when they execute a mission for the nation. SWE leaders focus on preparing individual ships for these potential missions. For the SWE, the final measure of performance is a “warship ready for tasking” across multiple possible missions, an output with a quality measure (Robinson, 2007, p. 28; Hatry, 2001). The SWE designed its performance management system to support the process of making ships ready and it is the expectation of Navy leaders (which we refute) that the system can also drive the budgeting process.²

¹ The leadership behind the initiative seeks to improve the cost-effectiveness of implementing the nation’s maritime strategy. See www.navyenterprise.navy.mil.

² Although Navy leaders express their desire to “drive” the budgeting process, we recognize that past information only informs the

While a component of a navy may seem like an unusual subject for a case study if one hopes to generalize findings, there are attributes of this case germane to many public organizations that struggle with performance based management and budgeting. The surface force provides an outcome (readiness) that is difficult to define and measure like other societal goals such as justice or public health. It provides an outcome whose causal factors are not clearly understood, like crime or poverty. Work processes have both routine and non-routine components conditioned by externalities. The SWE depends on the support and cooperation of other organizations to attain desired outcomes. Its functions and levels of resources are determined by political processes and not solely through rational management. Given the drive towards performance-based budgeting, this study is not only timely but of interest to practitioners and policymakers alike.

In the next section, we review the performance management and performance budgeting literatures relevant to our case study. We then describe the SWE and explain the performance management system and how leaders measure ship readiness. In the fourth section, we examine in greater detail this case's critical issues of performance measurement and cost. We explain some of the benefits and shortfalls of measuring readiness using the SWE's chosen algorithms and resulting performance indicators. We also discuss problems in aggregating the measures to get at overall ship and SWE effectiveness. We then discuss how the SWE uses cost analysis, some of the difficulties in measuring costs of inputs used to generate readiness indicators and why that is problematic for budgeting. After that, we present our findings and results, grounded in the literatures, and make recommendations for the SWE. Finally, we con-

budget process; that is, past performance does not dictate future decisions, but informs them.

clude and make recommendations for public organizations who seek to use performance systems to inform management and budgeting.

2. LITERATURE REVIEW

We derive our diagnostic framework from both the performance management and the performance budgeting literatures as those literatures apply to public sector, service-oriented organizations.

1. *Performance Management*

Robinson (2007, p. xxvi) defines performance management as the broad and systemic use of formal information to improve public sector performance, especially in the areas of human resource management, strategic planning and budgeting. Program budgeting is a mechanism for using performance information to influence priorities in resource allocation decisions (Robinson, 2007, p. 48). The U.S. military uses a program budget that classifies transactions into activities and programs. These activities and programs relate to and implement policy objectives. Ideally, leaders assess those activities and programs and measure their performance against objective criteria. With respect to specific activities and programs conducted by an organization, performance management systems measure and evaluate inputs to activities, or work to outputs (efficiency), and outputs to outcomes (effectiveness).³

Frumpinkin and Galaskiewicz (2004) and Robinson (2007), among others, note that government organizations have the least direct control over inputs and the least precise indicator of outputs of any type of organization. Performance management is often hampered by

³ We use “efficiency” and “effectiveness” as the public administration literature does. See, for example, HM Treasury (2001).

the lack of control of the quality and quantity of some inputs and the difficulty in finding appropriate output measures. Ambiguous causal relationships, environmental contingencies, and lag times contribute to the uncertain link between the production of outputs and attainment of outcomes (Havens H., 1983; Heinrich, 2004). In the case of the provision of public services, good outcome measures are problematic. Keeney and Gregory (2005) state that measures of objectives should be unambiguous, comprehensive, direct, operational, and understandable. Grizzle (1985) provides a consistent list of desirable attributes in her work on performance budgeting.

When attempting to bridge from performance management to performance budgeting, cost per unit of something (input, activity, output or outcome) is a primary consideration. Generally, activity-based costing uses input budget data (costs) to connect specific activities to outputs to support management decisions (Brown, Myring, and Gard, 1999; Mullins and Zorn, 1999; Williams and Melhuish, 1999). (Euske, Frause, Peck, Rosenstiel, and Schreck, 1999, p. 9) provide guidance on applying activity-based costing to service processes; they suggest tracking inputs and their resources relative to the output (service) the customer expects, “balancing that perspective with how to manage the service within the enterprise”. Such suggestions seem obvious in principle but are difficult in practice given the difficulty in defining outputs and, as the public budgeting literature shows, limitations of public spending data. Smith (2007) notes the specific difficulties in valuing national defense outcomes.

2. Performance budgeting

Robinson (2007, p. 1) suggests budgeting is the financial component of performance management, broadly referring to financial processes designed to “strengthen the linkage between funding and results” using

information in the performance management systems. Efficiency in performance budgeting has both an allocative component (results achieved through public expenditures) and a technical one (the cost of achieving the results). He further notes that performance budgeting can take different forms depending on the goals of the organization: some use it to improve spending prioritization or to emphasize program technical efficiency; some use it to fund future expected results or strengthen the understanding of the link between past results and spending decisions in order to affect future budgets; some create managerial incentives and others do not; and some emphasize outputs where others emphasize outcomes (Robinson, 2007, p. 15).

Havens (1983) notes the difficulty of integrating performance information into the budget process, specifically citing three impediments. First, offices that evaluate performance are often organizationally distinct from resource allocation offices. Second, the budget process and the evaluation process operate on different perceptions of time: budgeting is calendar-driven and evaluation is often event-driven. Third, budget analysts and program evaluators employ different analytical frameworks.

Empirical evidence suggests that the U.S. federal government, many state governments and other countries use performance information in the management of programs and display the information in their budgets; however, there is little evidence that spending decisions are greatly influenced by the performance information (Schick, 2002; Melkers and Willoughby, 1998; Jordan and Hackbart, 1999; Congressional Budget Office, 1993). Basing the budget on performance may be an unrealistic objective and performance information should only be expected to inform the budget process (Joyce, 1993; Schick, 2007). Flury and Schedler (2006) note the difficulties in serving both political and managerial needs with performance budget data. The produc-

tion and processing of information by the various actors in the budget process are such that it is unrealistic to assume budgets can have a pure performance basis.

Lu (1998) notes that performance budgeting has evolved from simple input and output measures to measures of efficiency and program effectiveness, but that the success of such systems hinges on the quality of measures (addressed above) and acceptance by decision-makers. Grizzle (1987) also notes that properly constructed incentives for managers and budgeters must be aligned with performance information. Sub-optimal behavior can result from mismanaging both actions and resources according to separate performance indicators, and sub-optimal behavior may occur at different levels of an organization. Managers may not want to be held accountable for outcome measures that have elements beyond their control. Organizational practices create incentives to manage performance, but disincentives to be accountable through the budget process – showing efficiencies currently takes funds away from efficient organizations (“use-it-or-lose-it” (Niskanen, 1971)) whether they are effective or not. Schick (2008, p. 8) so accurately comments:

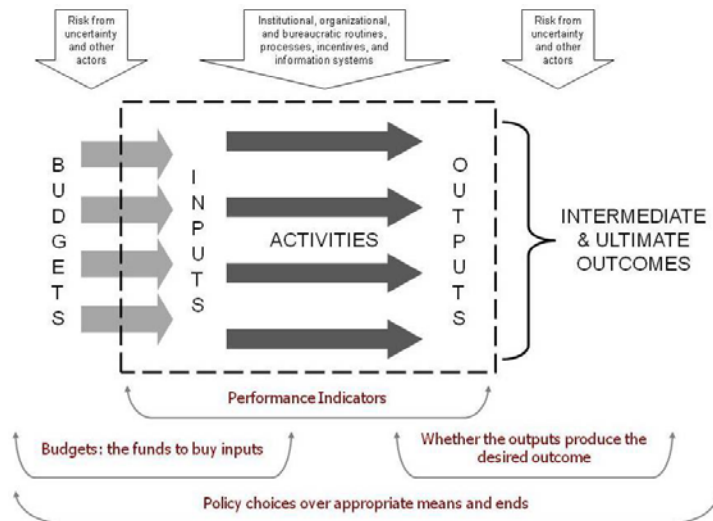
The ‘agency’ problem is especially acute on matters of performance, because adverse results can prejudice an entity’s budget. A resourceful manager once explained his behavior: ‘P[erformance] B[udgeting] requires me to load the gun that will be pointed at my head; as a manager, it is not hard for me to disarm the gun.’

Furthermore McNab and Melese (2003, p. 77) note that the traditional government control budget exists primarily to insure accountability and support appropriations processes, not to improve performance.

Integrating these literatures, we developed the following graphical depiction of the components of and relationship between a performance management sys-

tem and a performance budgeting system (Candrea and Webb, 2008). The left side shows the budget authorities by appropriation or line item that purchase inputs which, through a set of activities, convert to outputs. The outputs then combine to produce intermediate or ultimate outcomes. The dotted line inside the figure represents the boundary of the performance management system where managers concentrate on efficient production functions. The budgeting system operates outside the dotted box by validating the outputs and outcomes as a desired policy objective and by providing the budget authority to implement that policy. It is an open system, affected by the environment. Institutional, organizational, and bureaucratic routines, processes, incentives, and information systems (including financial accounting systems) affect the effectiveness of the system. It is through this framework we diagnosed the case of the Surface Warfare Enterprise.

Figure 1.
Performance Management and Budgeting Framework



3. NAVY SURFACE FORCES AND PERFORMANCE MANAGEMENT

The Navy, like all the military departments, provides assets ready to deploy in defense of the country. It provides personnel and trains and equips these resources, having them ready to support military operations conducted by the combatant commanders, key military leaders who manage a regional or functional area. Many organizations within the military services use performance management systems and attempt to inform the budget process using them; this paper focuses on a part of the shore component of the Navy, the one responsible for supporting ships. The shore component is organized into three "type commands," responsible for the military readiness of specific types of assets: aircraft, surface ships, and submarines. We focus our research on the surface force (SURFOR), under the command of a 3-star admiral, which currently supports the 162 surface ships of the U.S. Pacific and Atlantic Fleets.⁴ SURFOR manages approximately \$5.2 billion in annual operation and maintenance funds for the readiness of the surface fleet.

The Surface Warfare Enterprise (SWE) was established in 2005 under the auspices of the Navy Enterprise initiative. The SWE is an organizational construct that seeks to integrate the efforts of the Navy headquarters branch responsible for surface warfare policy (including approval of budgets); the Naval Sea Systems Command, responsible for designing and building ships (approximately \$9 billion annually in research and development funds and ship construction funds); and the SURFOR, responsible for active ship readiness. As part of the SWE, SURFOR seeks to optimize warfighting readiness of the Navy's surface fleet. Navy leaders be-

⁴ Ships are based in San Diego; Pearl Harbor; Norfolk; Mayport, FL; Ingleside, Texas; Everett and Bremerton, Wash.; Bahrain; Yokosuka, and Sasebo, Japan. (Navy Times, 2008).

lieve continuous process improvement (technical efficiency) in the core areas of maintenance, modernization, logistics, manning and training will create budget slack so the Navy can buy more ships, ammunition, and fuel (allocative efficiency).

1. *Matrix organization*

The SWE prompted SURFOR headquarters to reorganize along the lines of a matrix organization with functional and product line managers. Functional managers mirror the performance management system based on five critical performance algorithms or “figures of merit.” These correspond to personnel, equipment, supplies, training, and ordnance, or the acronym PESTO. Each functional manager oversees his respective PESTO area across all ship types. That is, there is a senior officer in charge of personnel, another in charge of equipment maintenance, and so on, who manage those matters for all ships.

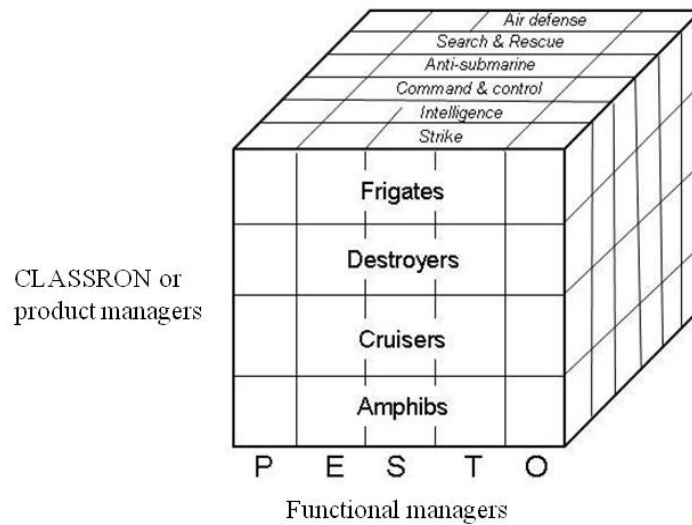
Product line managers, on the other hand, are responsible for all PESTO areas for a given ship type. Called class squadrons (CLASSRONs) and led by an officer of equivalent military rank to the functional managers, they are responsible for the overall readiness of one of four types of ship: frigate, destroyer, cruiser and amphibious.⁵ Each class of ships has unique systems, requirements and capabilities. SURFOR must prepare individual ships according to the ship’s technology and expected mission requirements.

To meet the Navy’s goal to project power anytime, anywhere, ships must be ready to function independent-

⁵ We note here that CLASSRONs are a SURFOR organizational element used to manage the preparation of ships for deployment and do not replace the operational chain of command that includes similarly titled organizational elements, such as destroyer squadrons (DESRONs).

ly and interdependently, complemented by advanced technological reach from other assets. Thus, navy ships are first evaluated for mission readiness independently, which is the proxy for output, and the ships are evaluated again by the combatant commander (at some point) within the group of assets with which it deploys. This second evaluation is outside the scope of the SWE's initial responsibility to provide a ready ship. The belief inherent in the system is that a properly trained and assessed individual ship will be capable of successfully integrating with others for all possible missions. Candreva and Webb (2008) created Figure 2 to show the matrix relationship among missions, ship type, and readiness indicators.

Figure 2.
Missions, ship type, and readiness indicators



Ship readiness is measured and reported in a fashion that is consistent with the overarching Defense Readiness Reporting System (DRRS), a defense-wide system for reporting military unit readiness for a given mission. Missions are comprised of discrete mission-essential tasks. A ship preparing for an anti-submarine

mission, for example, may be expected to perform tasks such as evading, detecting, tracking or engaging a submarine. Functional managers evaluate the ability to perform each task according to the five performance indicators (PESTO): sufficiently trained people, requisite equipment and weapons systems in proper working order with sufficient logistics support.

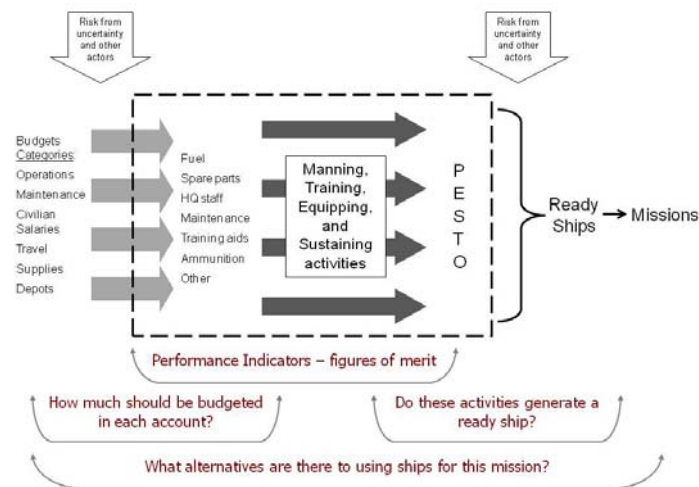
2. The SWE's Performance Management and Budgeting Framework

Figure 3, SWE Performance Framework, depicts the relationships among budget authority, inputs, outputs and outcomes for the surface navy (Candrea and Webb, 2008). Budget authority derives from various congressional appropriations justified by broad mission statements, detailed objects of expense (salaries, travel, utilities, supplies, rent), and longstanding performance measures that differ from the newer PESTO measures. The formal budget documents display input measures such as barrels of fuel and output measures such as underway days per calendar quarter, but say little about mission readiness. Once received, the appropriations fund the various inputs to activities that generate readiness as defined by PESTO, activities such as training, preventive or corrective maintenance, and operational exercises.

The inside of the figure, shown by a dotted line, represents the performance management system, where managers concentrate on efficiencies measured by the PESTO figures of merit. On the right side of Figure 3, outcomes are ships ready for tasking for different missions. From the perspective of a combatant commander, who ultimately decides what assets to employ and whether the mission was effective, ready ships are an input. Indeed, an argument can be made that a ready ship is actually an intermediate outcome to the larger defense mission. In this study, we correlate PESTO indicators to five proxy levels, each corresponding to the

quality of activities taken to measure ship readiness. Taken together and with human interpretation, they provide an overall picture of a particular ship's availability to conduct a certain mission.

Figure 3. SWE Performance Framework



4. PESTO PERFORMANCE MEASURES AND COSTS

1. *Measuring an individual ship's readiness using the PESTO indicators*

On the inside of Figure 3, PESTO algorithms attempt to capture the relationships among the inputs, activities or processes, and outputs. Replacing the "metric mania" (where the sheer number and disorganization of metrics makes evaluating, comprehension and accountability problematic (Casey, Peck, Webb, and Quast, 2008), at the SURFOR leadership level, PESTO attempts to simplify performance measurement. PESTO indicators are proxies, standardized along a 0-100 scale, and assigned "green," "blue," "yellow" and "red" by scores of 90-100, 80-90, 70-80 and below 70, respectively. Each indicator proxies whether the ship can per-

form a certain type of mission relative to the functional contribution (personnel, equipment, etc.) and is an output measure negotiated within and agreed upon by SWE personnel. The maintenance performance indicator, for example, comes from an algorithm that assigns values to repair tasks weighted according to their impact on mission accomplishment. The personnel indicator captures both the quantity of sailors and their individual training and qualifications. Similarly, the training performance indicator derives from an algorithm that calculates the “right” training for the unit as a whole. Of the five performance algorithms, personnel, training, and maintenance are the most mature.

To illustrate the complexity of tracking inputs to outputs, consider the relationship between the personnel (P) and training (T) elements of PESTO. Management of personnel primarily focuses on the inputs, processes and outcomes related to ensuring a sailor with the requisite skills fills a particular job. Managers use measures of “fit” and “fill” to assess performance: fill measures the number of sailors assigned to a ship and fit measures the professional characteristics of those sailors. If, for example, a ship requires and has in its crew four navigators and there are four critical navigation skills but the four navigators collectively are certified as competent in only three of the skills, the ship is 100% full, but only 75% fit. Managers can correct this deficiency by training one of the sailors in the requisite skill or, in the course of the routine rotation of sailors to and from shipboard duty, identifying a sailor with the requisite skills to be the next assigned. Hence, one can see the interrelationship between the personnel management and training management functions.

Training comprises two components: individual and ship-level training. Individual training may occur prior to a sailor’s arrival to the ship or it may occur once the sailor is part of the crew. The former is normally preferred because it increases the amount of time during

which the ship is ready to complete the various missions that sailor supports. If a sailor must leave the ship for training to become qualified in an area, her absence may reduce readiness in another area because each sailor supports multiple mission areas. Those who manage personnel and training readiness monitor the continuous process of sailor assignments, initial qualifications, gaps between current and desired states, and training events.

It is not enough to populate a ship with sailors with requisite skills: the sailors must demonstrate the capability to work together, employing the ship's technology, in a manner that assures their ability to meet mission requirements. Thus, managers measure ship level training in terms of the percentage of mission areas a ship has been certified as able to perform, the time it takes a ship to complete the certification process, and the cost associated with the certification events.

Taken together, these two functional areas provide other useful management information. For instance, navy leaders determined that a 90-95% fit measure is a reasonable level to expect given the system complexities of recruiting, training, assigning and retaining sailors, but ships can generally perform well if they are manned at 103% fill. The few extra people adequately compensate for the missing skills. The system, however, is far from comprehensive. The training management system, for instance, is not adequately linked to the maintenance management system. Many maintenance tasks are event-driven (e.g., each time a gun is fired for training, several preventive maintenance tasks must be performed) but those maintenance costs are not part of the training cost computation. Further, the sailors' salaries are centrally managed by the Navy, not by the SWE, so the fit-fill trade-off is miscalculated and may lead to a suboptimal decision.

2. Measuring readiness and outcomes for the SWE as an organization

Despite individual usefulness, we found that managers cannot aggregate the PESTO performance indicators to their goal of a single measure of “warships ready for tasking.” It is not reasonable to aggregate stoplight scores. In some instances, a “good” indicator (green or blue) does not insure a ship can perform a certain type of mission. For example, a ship tasked to perform a search and rescue mission could be “green” for training, equipment, ordnance and maintenance, and could have nearly all personnel ready to go, but could be missing the one requisite swimmer needed to perform the rescue. Despite appearing “green,” the ship cannot perform the mission and is not ready. The one missing item can cause the entire readiness indicator to be “redlined,” or dropped from a readiness status. By contrast, a ship might be at a lower-than-green level due to several minor problems that cause the algorithms to drop its scores, but is still be able to perform the mission. In another case, the commanding officer might feel ready to perform a certain type of mission because of an innovative work-around, and thus judge his ship as “ready” in spite of the measure. Finally, the notion of “warships ready for tasking” begs the question, “ready for *what* tasking?” A fully capable warship may not be necessary or prudent; for example, a ship assigned to a humanitarian assistance mission need not be concerned with a degradation of anti-submarine warfare capabilities. While this is sometimes considered *ex post* (the ship that is not anti-submarine warfare ready becomes the one assigned to the humanitarian mission), cultural norms that favor full mission readiness at all times do not consider it *ex ante*.

We find the PESTO scores are individually useful for directing action at the functional and product line management levels and for aggregating resources to be used at the margin (e.g., funding the highest priority maintenance repair – perhaps from a redlined ship – or sending a sailor to a training course). Separate perfor-

mance indicators can result in sub-optimal behavior, though. The indicators defining a specific mission area for a ship type can meaningfully drive actions, but aggregation across the third dimension is problematic. To get a clear understanding of overall “effectiveness” at the SWE level requires leaders to interpret the scores, reading written documentation supporting the scores, and asking questions when necessary. A clear understanding of effectiveness at the mission level – the effectiveness of the ship in performing the mission, is out of the scope of the SWE’s measurement system. However, leaders receive information that can be fed back into their system about strengths and weaknesses in their preparation of a specific ship for combat operations.

Lastly, occasionally the SWE overrides the performance information to drive specific organizational behavior. As a ship enters a maintenance period, repair tasks should be prioritized such that the equipment (maintenance) scores rise in the most critical mission areas. Occasionally, a particular ship modification or repair is considered a high priority but does not link to a specific mission area. One example given was the conversion of a restroom on an all-male ship to accommodate the addition of female sailors. Because this task was mandatory it was assigned a PESTO score of 100, meaning the ship scored a zero until the conversion was complete. Doing so fundamentally changed the purpose of the performance management system from an evaluation system to a control system (Behn, 2003) and masked the effects of all other repairs.

3. Costs and Budgets

Typical of many public sector performance budgeting attempts, the SWE has yet to make the leap from its longstanding encumbrance-based budgeting and accounting systems to a system of cost accounting that will provide adequate performance-based cost informa-

tion (Pollitt, 2001; Berman and Wang, 2000; Evans and Bellamy, 1995). The challenge is not technological, but cultural. Despite rhetoric of cost-consciousness, the SWE's primary financial concern is to obtain sufficient appropriations to operate and maintain the fleet at peak readiness and the secondary concern is to meet fiduciary responsibilities. The term "cost" is used synonymously with "obligation" even though there are important differences. An obligation is recorded at the time a contract is awarded or repair part requisitioned to meet the fiduciary responsibility of accounting for appropriated budget authority. It denotes that there is less budget remaining to spend, but it does not indicate the economic event normally associated with cost, the consumption of an input. That event often occurs much later. The prevailing belief in the SWE has been that the more one obligated, the more something cost; and, despite their stated intent, such beliefs drive analysis and information gathering today.

The navy at large does not have a cost accounting system of the type managerial accountants in the private sector might expect to find in a large organization. The financial systems that exist support the appropriation-based fiduciary responsibility of managers, and data are largely limited to obligations on objects of expense by organizational units within fiscal years. The navy does not well link financial data to processes or outputs, and non-financial information systems that support processes that consume financial resources are not designed to provide adequate cost data. For example, maintenance systems manage repair activities and the obligation of funds for repair parts may be included, but not the costs of labor, indirect materials, and allocated overhead.

We find five types of cost analysis in practice in the SWE. In the first type, analysts mine data to determine what is being purchased and to assess whether those purchases could be reduced. Such studies have shown

that grey paint is the single most frequently purchased item, which led the SWE to examine lower-cost alternatives to traditional paint. Analysts have also shown particular repair parts are ordered more frequently than expected leading to cost-benefit analyses of re-engineering the component. These analyses support idiosyncratic technical efficiency efforts, but do not support attempts to allocate efficiently.

Secondly, spending by ships of the same class is compared based on homeport, or whether ships are assigned to the Atlantic or Pacific force. Such comparisons may yield information about differing regional maintenance or training practices, which can be helpful management information. Often, however, such comparisons lead to less productive discussions of fairness and equity in the distribution of resources.

In the third type of cost analysis, the SWE has built a system of “bridgeplots” in navy parlance, or what might be called “dashboards.” Analysts chart cumulative year-to-date spending against rolling averages of performance. The mismatched time scales are difficult to interpret and spending starts at zero at the start of a fiscal year. Managers who have historically cared more about managing appropriations than cost understand the spending plot; however, it is literally impossible to see the relationship between spending and performance measures.

In the fourth type of cost analysis, the SWE uses the stoplight-coding schema for readiness indicators and attempts to compute the cost to move a ship from one (stoplight) status to the next. SWE leaders intend to allocate funds to gain maximum benefit in terms of readiness. Two problems exist with this analysis. First, given the limitations of the accounting systems and knowledge of causal relationships, leaders have little confidence in the amount needed to move a ship from one level to the next. Second, even if analysts well un-

derstand costs, the stoplight system encourages suboptimal decision-making, as resources tend to flow to the ships just below a threshold to give the appearance of progress, even if there are more important problems on other ships.

Finally, analysts assign spending to missions in an attempt to understand or manage the cost of those missions. This is an admirable attempt to link cost to readiness, but there are problems with the method. First, the mobility mission (the ability of the ship to simply move from one location to another) accounts for nearly half of the funds spent because of high fuel consumption. Mobility missions include things such as propulsion and electricity generation, fundamental to all other missions. Thus, it should be viewed not as a “product center” but a “cost center” that provides basic services to other missions to which the mobility costs should be allocated on some logical basis. Second, as noted previously, the SWE considers a cost has been incurred when something is requisitioned, not when it is consumed and there can be significant lag times between the events. Parts may be consumed that were requisitioned years ago and just now taken off storeroom shelves. Third, two significant factors affecting readiness, the salaries of the sailors on the ships and the original construction and capital improvements to ships, are not included in the SWE’s cost assignments because SURFOR does not control those funding lines. To the SURFOR, those costs are not relevant to the decisions they make and are ignored (Robinson, 2007, pp. 55-56); to the SWE, they are relevant, but are not yet systematically captured in the performance management or budgeting systems. This is not surprising for political, managerial, and technical reasons. Politically, control of financial resources is a significant source of institutional power and managers share information reluctantly (Salancik, 2003). Policy makers and managers make different types of decisions that rely on different types of information (Flury and Schedler, 2006). Technically, public

sector accounting struggles with the notion of depreciation expenses and allocating the costs of fixed assets to products and services. Unlike the for-profit model where the matching principle of accounting states that all expenses must be matched to the revenue they generate and capital assets are assumed to generate revenue, links between revenue and capital expenditures in the public sector are normally confined to the realm of financing. Government's aim is not to generate revenue, it is to provide services, so both the logic and practice of accounting for capital assets is problematic (Chan, 2003; GASB, 2006).

While each of those forms of cost analysis yields some specific benefits, there has not been a systematic use of the analyses in the formal Navy budgeting process. Neither the process for budgeting for ship operations nor the format and content of those budgets has changed appreciably. The performance management system operates within the CLASSRONs and SURFOR at locations geographically and organizationally removed from the headquarters where budget allocation decisions are made. The SWE is just beginning to use these performance data as part of the justification for budget requests, but the use or acceptance of the data are not routine. SWE leaders find that the competition for resources has not fundamentally changed. In the end, budget allocations remain a political choice between desired policy outcomes and the existence of the performance data may lend a little more veracity to the requested sum, but it does not change the nature of the decision-making process. The SWE recognizes the need to augment their empirical "facts" with political suasion.

5. FINDINGS

We find the SWE's performance management system is logical, detailed, comprehensive and reflects organizational structures and management practices.

While each critical performance measure or PESTO function is incorporated into the performance management system through tasks and linked to mission areas, interrelationships among them are not well defined. Aggregation to a single measure of readiness across missions is not meaningful in this context. More detailed performance data are meaningful and managerially useful; the organization uses the data to improve technical efficiency but in a manner that may sub-optimize allocative efficiency. There is not a logical link between the creation of budget slack by improving a given process and an efficient allocation of that slack. The SWE has a system for the first, but not necessarily for the second.

Institutional norms and overarching concerns about preserving, consuming, and expanding appropriations result in the use of inappropriate proxies for cost. Cost analysis, then, is compromised. It may result in technical efficiency gains, but does not support the goal of allocative efficiency. Some vital resources used in producing readiness come from outside the SWE, are funded in other organizations' budgets, and are not considered in the performance management system. Occasionally, managers co-opt the system to induce management behavior the system was not designed to induce. Taken together, we find SWE managers have established a system that produces useful managerial information about specific functions and tasks and that they are generally enthusiastic about the system because of its mission focus. We also find the system cannot measure efficiency of the full production function to get a ship ready and cannot support the budget process by providing a strong link between a level of funding and level of fleet readiness. We also observe a performance management system designed around the factors the SURFOR can control more than one designed to capture the full set of variables affecting fleet readiness. Sailor salaries and ship construction and modification costs are borne by other organizations from other appropriations

and are not factored into the cost of readiness. The cost of filling the ship to 103% manning levels is not captured, but does have implications for broader navy-wide budgeting.

Consistent with the literature, the SWE has found it difficult to create a useful outcome measure, but has built a performance management system that provides detailed information about specific outputs. SWE leaders would argue they have a single precise outcome measure (warships ready for tasking), but its utility is questionable. They would do well to abandon the idea of a single measure and accept that a handful of measures is more descriptive and provides more management information. For the most part, the SWE's technical measures are unambiguous, direct, operational and understandable and they are well documented. Their corresponding financial measures are more ambiguous and less direct and should better correlate with the types of management decisions the SURFOR confronts. Appropriate cost data should be used when seeking to improve the efficiency of processes; obligation data should be used when deciding how to spend appropriated funds. Financial systems should be developed to generate both consistently.⁶ Further, the cost analyses do not often separate price effects from efficiency effects – the cost of an input, like paint, is often measured in dollars rather than gallons. Controlling for price variation is necessary in order to isolate and understand the amount of an input to a process. Currently, managers cannot use the system to fully assess how much of an apparent change in efficiency is due to changes in the price of

⁶ Many organizations combine planning, manufacturing, distribution, shipping, and accounting systems into an enterprise resource planning (ERP) system that integrates all of these functions into a single system. An ERP is designed to serve the needs of each different organization within the enterprise. An ERP is in development and deployment in the U.S. Navy, but has not been deployed in the SWE and will not for some years. Thus, it was not part of the case analysis.

inputs, substitution of inputs, improvements in technology, better training, or other productivity effect.

Further complicating the connection between budgets and performance is that budgeting operates on a cyclical timeline while program performance is continuous. At any moment, four budget cycles occur simultaneously: one budget is being executed as the next is being enacted, a third is being formulated, and the requirements for a fourth are under study. Even with robust accounting systems, establishing the link between readiness of the ships, the consumption of inputs that generated that readiness, and a specific year's budget is empirically complex. What expenditures lead to a particular ship being ready for a particular mission? The spare parts in the ship's storeroom may have been purchased days or years ago. The training of the sailors may have occurred weeks or months ago. The collective experience of the captain and crew, not to mention the ship itself, may be two decades old. By extension, to which budget cycle (i.e., which fiscal year's level of funding) does one ascribe a requirement for funds to purchase a set of inputs that will be used immediately, later, or perhaps never? The reality is that such decisions are often made in order to consume expiring appropriations or to meet another fiduciary threshold and not because the performance management system recommends it.

As Havens (1983) notes, program analysts and budget analysts employ different analytical frameworks. The performance management system is designed by war fighters and aligns functional contributions to missions they support. The budget, on the other hand, is designed by technocrats and aligns objects-of-expense into a program structure. Short of redesigning the format of the budget, budget analysts are unable to process the performance information. For now, the broader Department of the Navy budget is formulated in the Pentagon, away from the surface force, by analysts and bud-

geters who do not and cannot process the detailed information contained in the performance management system. Program analysts who determine desired future budget levels and types of activities take a more strategic view of the SWE than managers who operate the fleet. If program analysts foresee a future different from the past then even optimal data provided by the fleet may not help determine future budgeting requirements. If budgeters weigh the need for submarines and aircraft against the need for surface ships, no common basis for comparison exists. When weighing whether cruise missiles from a surface ship or smart bombs from an Air Force plane should perform a strike mission, the performance management system does not help.

Further confounding this organizational separation is the fact that many of the SWE's assets serve multiple purposes. A destroyer, for instance, can provide air defense support, anti-submarine support, and can strike targets at long distances with cruise missiles. Program analysts aligned by mission areas will incorporate the destroyer in their plans. A cost accounting system could assign a percentage of the ship's cost to each mission area, but one cannot budget for a fractional ship. All the resources for that ship are necessarily assigned to one of those mission areas, overstating its cost and understating the cost of other areas.

The combatant commanders who employ the ships in pursuit of national security objectives are also organizationally distinct from the budget process. By design, the operational commander's attention is devoted to current operations in a theater while the navy staff in the Pentagon devotes its attention to building and supporting the navy. Those who can best assess the value of the military asset have no voice in the budget process. As Smith (2007) suggests, it may be preferable to value ships on military terms than economic ones.

Finally, as Joyce (2003) reminds us, the best performance management system cannot drive budgets. In the case of the SWE, the system (and the information it contains) is conceptually and organizationally distinct from the responsibility and accountability for determining the budget. The SWE has not built a bridge between the performance management system and the budgeting system. Given Schick's (2007, p. 16) belief that "[p]ublic organizations would do well to deploy performance budgeting as an analytic tool because few have the capacity to ground budget decisions on it," we recommend SWE leaders think less about how to determine budgets through their system and more about how to use performance information to influence budgets. It may not be possible to answer, "how much fleet readiness does a certain amount of funding provide?" but it is possible to quantify the impact of incremental adjustments to budgets. Joyce (2003) suggests that performance information should inform budgeting. Navy leaders who are frustrated that budgeting is not easier or more automatic should accept that easier and automatic are not reasonable expectations.

6. CONCLUSIONS AND RECOMMENDATIONS FOR PUBLIC ORGANIZATIONS

This study suggests that organizations implement performance management and performance budgeting systems for different reasons, and the form taken by the systems should match desired outcomes from the use of the systems. Managers and leaders should construct proper measures and align management incentives with desired outcomes and behavior; otherwise, individual parts of the organization will sub optimize. This research confirms that trying to do too much with a given system results in inappropriate use of data, mismanagement of resources or misalignments of actions. It confirms that a desire for a single measure of performance or effectiveness is, to use Hatry's language, "fashionable but should be looked at with considerable

skepticism” (2002, p. 352). And it confirms that precision does not equal accuracy. Managers of hard-to-define outcomes and complex production functions should be content with generating a set of measures that captures the complexity of the issues.

The research also confirms that those in an organization who evaluate and act on performance measures operate separately from those who allocate resources. Budget analysts and program evaluators employ different analytical frameworks, and the budget process and evaluation process operate on different perceptions of time. Thus, using good data to inform the budget process may be the best leaders can hope to achieve from their performance management system.

Finally, an organization’s managers must recognize that policymakers make different types of decisions from their own, and need different information. This study illustrates the usefulness of performance information to an organization’s managers and the frustrations they experience trying to use that information to influence policy and funding. In recognition of these points, the SURFOR appointed a civilian deputy to the admiral to influence policymakers and other SWE organizations; however, it is too early to tell whether a “bridge maker” can help improve use of performance data in budgeting. Future studies should assess the effectiveness of such a “bridging” mechanism from performance to budgets.

In sum, this study provides some diagnosis of and suggestions for better using performance management systems to increase the efficiency and effectiveness of public sector organizations and provides evidence of the challenges faced when attempting to use that information in public budgeting.

REFERENCES

- Behn, R. D. (2003) "Why Measure Performance?", *Public Administration Review*, 63(5), pp. 586-606.
- Berman, E., and X. Wang (2000) "Performance Measurement in U.S. Counties: Capacity for Reform", *Public Administration Review*, 60(5), pp. 409-420.
- Brown, R., M. Myring and C. Gard (1999) "Activity-based Costing in Government: Possibilities and Pitfalls", *Public Budgeting & Finance*, 19(2), pp. 3-21.
- Candrea, P. J., and N. Webb (2008) *Exploring the Link between Performance and Resource Allocation in the Navy Enterprise*, Center for Defense Management Research, Monterey.
- Casey, W., W. Peck, N. Webb, and P. Quast (2008). "Are we driving strategic results or metric mania? Evaluating performance in the public sector," *International Public Management Review*, 9(2), pp. 90-106.
- Chan, J. L. (2003) "Government Accounting: An Assessment of Theory, Purposes and Standards", *Public Money and Management*, 23(1), pp. 13-20.
- Congressional Budget Office. (1993). *Using performance measures in the federal budget process*, Government Printing Office, Washington.
- Euske, K. J., N. Frause, T. Peck, B. Rosenstiel, and S. Schreck (1999) "Applying Activity-based Performance Measures to Service Processes: Process Relationship Maps and Process Analysis", *International Journal of Strategic Cost Management*, Summer, pp. 3-16.
- Evans, P. and S. Bellamy (1995) "Performance Evaluation in the Australian Public Sector: The Role of

Management and Cost Accounting Control Systems”, *International Journal of Public Sector Management*, 8(6), pp. 30-38.

Flury, R. and K. Schedler (2006) “Political Versus Managerial Use of Cost and Performance Accounting”, *Public Money & Management*, 26(4), pp. 229-234.

Frumpinkin, P. and J. Galaskiewicz (2004) “Institutional Isomorphism and Public Sector Organizations”, *Journal of Public Administration Research and Theory*, 14(3), pp. 283-307.

Government Accounting Standards Board (GASB) (2006) *Why Government Accounting and Financial Reporting Is - And Should Be – Different*, Retrieved March 20, 2010 from http://www.gasb.org/white_paper_mar_2006.html

Grizzle, G. A. (1987) “Linking Performance to Funding Decisions: What is the Budgeter's Role?”, *Public Productivity and Management Review*, 41, pp. 33-44.

Grizzle, G. A. (1985) “Performance Measures for Budget Justifications: Developing a Selection Strategy”, *Public Productivity and Management Review*, 39, pp. 328-341.

Hatry, H. P. (2002) “Performance Measurement: Fashions and Fallacies”, *Public Performance & Management Review*, 25(4), pp.352-358.

Hatry, H. (2001) “What Types of Performance Information Should be Tracked?”, In D. Forsythe, *Quicker, Better, Cheaper?: Managing Performance in American Government* (pp. 17-34) Rockefeller Institute, Albany.

Havens, H. (1983) "Integrating Evaluation and Budgeting", *Public Budgeting & Finance*, 3(2), pp. 102-113.

Heinrich, C. J. (2004) "Improving Public Sector Performance Management: One Step Forward, Two Steps Back?", *Public Finance and Management*, 4(3), pp. 317-351.

HM Treasury (2001) *Choosing the Right Fabric: A Framework for Performance Information*, HM Treasury National Audit Office, London.

Jordan, M. M. and M. M. Hackbart (1999) "Performance Budgeting and Performance Funding in the States: A States Assessment", *Public Budgeting & Finance*, 19(1), pp. 68-88.

Joyce, P. G. (2003) *Linking Performance and Budgeting: Opportunities in the Federal Budget Process*, IBM Center for the Business of Government, Washington.

Joyce, P. G. (1993) "Using Performance Measures in the Federal Budget Process: Proposals and Prospects", *Public Budgeting & Finance*, 13(4), pp. 3-17.

Keeney, R. and R. S. Gregory (2005) "Selecting Attributes to Measure the Achievement of Objectives," *Operations Research*, 53(1), pp. 1-11.

Lu, H. (1998) "Performance Budgeting Resuscitated: Why is it Still Inviabile?", *Journal of Public Budgeting, Accounting & Financial Management*, 10(2), pp. 151-172.

McCaffery, J. and L. Jones (2002) "Assessing Options for Changing the Federal Government Budget Process", *Public Finance and Management*, 2(3), pp. 436-469.

McNab, R. M. and F. Melese (2003) "Implementing the GPRA: Examining the Prospects for Performance Budgeting in the Federal Government," *Public Budgeting and Finance*, 23(2), pp. 73-95.

Melkers, J. and K. Willoughby (1998) "The State of the States: Performance-Based Budgeting Requirements in 47 Out of 50", *Public Administration Review*, 58(1), pp. 66-73.

Mullins, D. and C. Zorn (1999) "Is Activity-Based Costing up to the Challenge When it Comes to Privatization of Local Government services?", *Public Budgeting & Finance*, 19(2), pp. 37-58.

Niskanen, W. A. (1971) *Bureaucracy and Representative Government*, Aldine-Atherton, Chicago.

Pfeffer, J. and G. R. Salancik (2003) *The External Control of Organizations: A Resource Dependence Perspective*, Stanford University Press, Stanford.

Pollitt, C. (2001) "Integrating Financial Management and Performance Management", *OECD Journal on Budgeting*, 1(2), pp. 7-37.

Robinson, M. (2007) "Cost Information", in M. Robinson (Ed.), *Performance Budgeting: Linking Funding and Results* (pp. 46-62), International Monetary Fund, New York.

Robinson, M. (2007) "Performance Budgeting Models and Mechanisms", in M. Robinson (Ed.), *Performance Budgeting: Linking Funding and Results* (pp. 1-21), International Monetary Fund, New York.

Schick, A. (2002) "Does Budgeting Have a Future?", *OECD Journal on Budgeting*, 2(2), pp. 7-48.

Schick, A. (2007) "Performance Budgeting and Accrual Budgeting: Decision Rules or Analytic Tools?", *OECD Journal on Budgeting*, 7(2), pp. 109-138.

Schick, A. (2008) *Getting Performance Budgeting to Perform*, Retrieved January 8, 2008, from World Bank: <http://siteresources.worldbank.org/MEXICOEXTN/Resources/ConceptPaperAllenSchickFinal.pdf>

Smith, R. (2007) "Valuing Defense", *Public Finance and Management*, 7(3), pp. 242-259.

Williams, C. A. and W. Melhuish (1999) "Is ABCM Destined for Success or Failure in the Federal Government?", *Public Budgeting & Finance*, 19(1), pp. 22-36.

ACKNOWLEDGEMENTS

This study was conducted with support from the Office of the Chief of Naval Operations (N09X), Navy Enterprise Office. The views contained in this paper are those of the author and do not necessarily reflect the position of the Department of the Navy. The authors appreciate the helpful comments of the anonymous reviewers. All errors are the authors' alone.

AUTHOR INFORMATION

Dr. Natalie J. Webb (njwebb@nps.edu) is Associate Professor of Economics at the Defense Resources Management Institute and a research associate with the Center for Defense Management Research in Monterey, California.

Philip J. Candreva (pjcandre@nps.edu) is Senior Lecturer of Budgeting at the Naval Postgraduate School and a senior associate with the Center for Defense Management Research in Monterey, California.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.