DOES THE CHOICE OF DESIGN IN GOVERNMENT PROGRAMS AFFECT PERFORMANCE?

Choosing Among Alternative Administrative and Budget Design Options for Improved Program Results

Stuart Kasdin

Trachtenberg School of Public Policy and Public Administration
George Washington University
skasdin@gwu.edu

Presented at the National Public Management Research Conference
June, 2011
“All human societies go through fads in which they temporarily either adopt practices of little use or else abandon practices of considerable use.” Jared Diamond

Government program administration is not a world of uniformly simple, lumbering bureaucracies. Instead, federal managers and Congressional authorizers have embraced a portfolio of administrative options, including outsourcing, block grants, competition schemes, and network structures (Hill and Lynn, Jr, 2005; Goerdel, 2006; Frederickson and Frederickson, 2006).¹ These program designs rely on a more decentralized and market-driven model, aimed at increased efficiency and improved results oriented by emphasizing accountability for outcomes, through managerial flexibility, decentralization, deregulation, and performance measurement. By placing control and responsibility in local managers’ hands, and avoiding the straightjacket of input controls and managerial rigidity, a new focus would be placed on outcomes. Organizational leadership could step back from top-down management by relying on accessible outcome-oriented performance measures and improved data.

Government operates more effectively when it focuses on outcomes, when leaders set clear and measurable goals, and when agencies use measurement to reinforce priorities, motivate action, and illuminate a path to improvement. This outcome-focused performance management approach has proved a powerful way to achieve large performance gains in other countries, several States, an increasing number of local governments, and a growing number of Federal programs…Outcome-focused performance management can transform the way government works, but its success is by no means assured. (OMB, 2010, p 73)

In the federal government, many of these reforms have been applied across all agencies and departments. The 1993 Government Performance and Results Act (GPRA) was applied to all programs across the federal government. The GPRA sought to enhance strategic planning and program management in the federal government.

Similarly, the U.S. Office of Management and Budget (OMB) Circular A-76, Performance of Commercial Activities, originally published in 1983 and last revised in 2003 requires

¹ Thanks to Eduardo Fernandez Maldonado for great research assistance.
competition between agencies and private sector suppliers for all government agencies, and applies widely to federal positions. These ideas of competition were emphasized in the Bush Administration’s 2001 President’s Management Agenda. This PMA included these and other NPA reforms that were universally applied.² It began with the premise that “Government should be: Citizen-centered, not bureaucracy-centered; Results-oriented; Market-based, actively promoting rather than stifling innovation through competition” (OMB, 2002, p. 4) It included initiatives in competitive sourcing (i.e., mandated competition between government agencies and external contractors), workforce planning and restructuring, expanded electronic government, and budget and performance integration.

However, one problem with universally applied reforms is that they are implemented without strategic overview and without deliberate weighing of the trade-offs and consequences for the individual programs.³ It is assumed that the reforms yield similar consistent benefits across program. That expectation may not be warranted. This paper considers whether the universal application of the various New Public Administration approaches is warranted, and when it is appropriate to apply the different tools.

The paper is organized as follows. First, I examine and categorize the different types of administrative designs. Then, I look at elements that influential in the success of those designs, the performance measures and the nature of the program oversight. By understanding the nature of the agency performance measures and oversight regime, I offer some theories as to when one should use different administrative and budget structure designs in order to get improved program results. I then empirically test these theories to see if I can predict the types of

² So too, the 1994 National Performance Review and the Reinventing Government efforts of the 1990s were applied across government.
³ Instead, the administrative reforms are implemented opportunistically, where agency and Congressional resistance is low, as opposed to where benefits are highest.
administrative and budget designs most likely to generate the highest performance, given better or worse performance measures designs.

**Alternative Program and Budget Designs**

There have been a variety of different approaches to categorizing and evaluating the choice of tools or program designs by government (Salamon, 2002; Howlett, 2002; Bemelmans-Viec, et al, 2003; Sandfort, 2007). These catalogue the advantages and disadvantages of different forms government programs – regulations, grants, loans, vouchers, direct expenditures, etc. – considering the program structures as the tools to accomplish program ends, as well as trying to develop dimensions or groupings (Peters, 2000). In Salamon’s (2002) work, the tools vary according to four dimensions of coerciveness, directness, automaticity, and visibility, which like Dahl and Lindlom (1953), are evaluated on a continuum basis. He assesses the impacts of these metrics, depending on the circumstances, with respect to effectiveness, efficiency, equity, manageability, and legitimacy. Dahl and Lindblom (1953)’s design classified instruments according to the nature of government influence (compulsion, persuasion, or information) and control (direct or indirect), instrument ownership (public or private), and autonomy (varying from government agencies to full autonomous entities), and on the nature of the instrument membership (voluntary or compulsory).

This analysis will build on these schemas, while attempting to simplify the analysis. The simple classification will divide government programs according to whether they rely on delegated authority or direct provision by federal agencies, and then the degree of controls or incentives placed on the program management. The former reflects the locus of program decisionmaking, while the latter reflects the controls or incentives imposed on that authority. See Table 1.
In the first quadrant are programs in which the principal – the government agency – relies on an agent to carry out the program functions. However, the authority is not unlimited. It is constrained, with controls on the agent limiting its discretion in making program choices. For example, EPA delegates authority to states for a variety of programs, such as solid waste and clean air, however, it must approve the state plans before authority is devolved. Even then, states must continue to carry out the approved plans fully or risk losing their authority.

There are some programs in which there are almost no limitations on what the fund recipient – usually states -- can do with funds received. The Department of the Interior’s National Park Service - Land and Water Conservation Fund State Grants provides matching grants for States and local governments to acquire and develop lands and facilities for public outdoor recreation, covering basketball courts and urban parks, as well as more environmentally important locations.

There are different degrees of fungibility associated with grants. Grants may be completely open-ended, like HUD’s Community Development Block Grant, allowing recipients (the states) to use funds for multiple missions or goals. The CDBG permits states to fund housing rehabilitation, neighborhood revitalization, public facilities, and job creation. Other types of grant programs’ authorities are more limiting, whether restricting the program to a small
set of individuals, without constraining eligible activities (e.g. HUD’s Native American Housing Block Grants) or controlling the services offered (HHS’ Office of Child Support Enforcement, which available to children by locating parents, establishing paternity, and establishing the obligations of parents to provide child support), but not necessarily the population receiving the benefits.

For quadrants 2 and 3, the degree of incentivization provided through a contract may differ based on contract length. When the US Forest Service offers a 5 year contract to management concessions in some national forests, there is significant competition available and hence, pressure on the contract winner to perform at the highest level (since the agency uses performance contracts). On the other hand, the Forest Service also leases out the management of dams and ski resorts to private firms for 50 year periods. The incentivization is low in those cases. Needless to say single source, no-bid, and/or cost-plus contracts do not offer strong motivation for good performance either.

The question of which of these design approaches to use is the subject of this paper.

The Quality of a Performance Measurement System

To evaluate the quality of a performance measurement system, we would start by looking at a program’s individual measures and consider their reliability and validity. That is, do the measures fluctuate because of errors and non-program reasons? Are measures evaluating what we think they should? Then, we would examine the program’s performance measures as a bundle: how does the set of measures fit together?

In the first case, a high quality performance measurement system will depend on how well we can measure the impact of an agency’s efforts in terms of impacting the program goals. Can the impact of environmental variation be controlled for? Can we separate agency efforts
from the environmental noise, including the efforts of other agencies (Rubenstein, et al, 2003)? Measures often contain a lot of “noise” so that the impact of work can be swamped by external factors and random error. To be meaningful, measures must also be reliable; i.e., offering good measurement accuracy.

One problem may be that unpredictable, uncontrollable external events may dominate a program’s performance. For example, exchange rate changes may swamp the impacts of any government program encouraging international exports. Likewise the performance of a program to help people get jobs can only be evaluated if we recognize that the state of the economy has the biggest impact on employment.

The other way that measures may not be reliable is if measurement error is common. Thus a water quality program would need to consider ambient temperature of the water in order avoid a measurement error that confuses the assessment of program impacts. Overall, performance measures with a lot of noise that dull the performance measure’s effectiveness. Noise means measures are less reliable to enforcing or encouraging agency behavior.

At a local level, many community services avoid these problems. Services like sanitation and recycling have ready counterparts in the private sector. Evaluating the performance of a public organization is easier when the work can be compared to others doing the same service. This allows one to assess the contributions of environmental influences on the outcomes measured and to compare the performance of different private or non-profit firms and public organizations.

For the federal government, however, program outputs may not be as frequently produced in the same form by the private sector. In the Government Performance and Results Act, programs uniformly assess their performance without comparison to external entities doing
the same task. That means that in assessing a program’s performance, an evaluator is forced to compare the performance at different points in time, as opposed to comparing performance to an external benchmark. Since economic and physical conditions (as well as appropriation amounts) vary at alternative points in time, the metric’s usefulness is limited.

The second area of concern for performance measure is that, even if agency performance measures are reliable, they may not be very faithful in reflecting the program goals accurately—i.e., the construct validity is low so that the measure(s) do not accurately capture the program’s mission. Is the measure assessing what we think? How well does the measure respond to the agency’s efforts; does it evaluate program contributions?

There are a number of common circumstances when this might be the case. For example, there may be multiple public and nongovernment programs with overlapping impacts, so that an individual program’s contribution is part of a collective contribution. In such a case the contributions of a single program become unclear and the performance measure may not indicate how well a program is doing. For example, programs supporting water quality improvement across a large watershed or targeting economic development in a poor region may have a difficult time distinguishing which program was responsible for any eventual changes.

Another way that a set of performance measures may not achieve high validity is when there is invariance in the short run or when an outcome is achieved only distantly in the future. Program measures with outcomes that take a long time to achieve, like a program to impact climate change may not be amenable for evaluating on-going performance, at least without intermediate measures as targets (Hatry, 2006). Finally, cases where programs are pursuing goals of deterrent behavior, in which success is measured by negative actions forgone, can be
difficult for performance measurement (OMB, 2003). Behavior that doesn’t happen is difficult to measure.

In addition, government programs are complex with multiple activity elements define the mission, plus there are multiple outcomes or missions. In some cases, this may not be a problem because the program’s the combination of measures are highly compatible. Typically, that occurs when the program has only a few performance measures and they are largely congruent; i.e., they do not measure missions that run counter to each other.

The opposite problem is non-congruent program goals. Agencies may have multiple program goals not all of which are fully consistent with the other. Sometimes to encourage one measure (e.g., large numbers of clients served or no waiting on the phone for any phone call) may mean other goals (e.g., extended and satisfied customer service) do not get fulfilled.

In an extreme case, an index of performance measures for the US Forest Service, which included both endangered species habitat protection and timber harvest levels, would not yield a meaningful result, since it balances environmental protection and economic development. Therefore, a program with a multiplicity of complex goals, especially where they do not reflect isomorphic goals (even if they are bundled in an index) will not generate high quality performance measurement.

Measuring a programs’ performance is difficult under the best of circumstances. The typical program’s array of complex, multidimensional traits, paired with imperfect information may not be easily evaluated using performance measures. Performance measurement quality can be improved when they are weak – performance measurement quality is not invariably an
essential characteristic of a program – however, it may always be difficult to achieve a high quality system in some cases.4

**Using Performance Measures as Part of Program Design or in Tool Selection**

We want to be able to evaluate when a program might delegate authority more fully to state or non-profit managers, with reduced external controls. Also, we would want to have an indication as to whether incorporating more incentives – emulating the competitive environment of a private sector is a good idea, and when we should rely more on traditional management approaches.

How do performance measures indicate which types of program design or structures we should use for a program?  Should program design change when the performance measures are complex, without external benchmark comparisons?  How should agency discretion vary as a result of the degree of clarity associated with program missions (p. 126, Ingraham, et al., 2003)?  Should this affect the determination on whether the government should aim to “row” or “steer” (Osborne and Gaebler, 1991)?

Within the private sector, the determination of whether a firm relies on internal production or market procurement of goods and services (“make it or buy it”) is based on transaction costs.  When asset specificity is high (i.e., fewer possible vendors to contract with), the risk of opportunism rises, and so internal production is preferred.  Also, as asset specificity

---

4 For example, Wilson (1989) uses the police departments as an example of “coping organizations” in which neither outputs nor outcomes are observable.  Neither the work activities of the police officers, like interactions on the street man, nor the outcomes, such as the degree of calm or order on a policeman's beat can be easily measured or observed.  However, in recent years, police departments have emphasized the use of police work for management by focusing on aggregates – what was measurable was the crime in an area or district. This could be tracked and compared to other districts. (Maguire and Uchida, 2000)
decreases, the competition-based benefits of market production outweigh the benefits of internal production (Williamson, 1985; 1999).

For the federal government, the level of transaction costs is reflected in the precision of the performance measures which determine the optimal type of incentives or delegation used.\textsuperscript{5}

Strong performance measures reflect the conditions for when Congress or OMB can attach incentives to the performance measures, steering the agency so that the incentives do not distort or misdirect agency management or leave out important goals. In general then, performance measure “quality” is defined as how the set of performance measures, tied to the basket of program goals, enhances the guidance of program direction, without goal distortions and encourages intensity of effort (Kasdin, 2010).\textsuperscript{6}

Accordingly, the ideal type of program design chosen for a program is dependent on performance measurement quality. Managing a program by relying on performance measures is ineffective when the quality of the measures is poor. As the reliability of the measures become less reliable, as the validity of measures decreases, the costs of using the measures for management decisions will rise. The key for categorizing programs as eligible for performance budgeting is based on the quality of their performance measures; specifically how ‘contractable’ they are.

Figure 1 shows the effect of a program’s contractability or performance measurement system quality (x-axis). The Y axis then reflects the outcomes of the agency’s work; the socially beneficial outcomes, reflecting the impact of the basket of performance measures and incentives

\textsuperscript{5} For the federal government, private markets are larger (potentially reflecting nation-wide pools) and so risks of opportunism and hold-up costs are lower. This may not be true when the government activity requires secrecy thereby limiting the pool of potential applicants from the private sector. Hold up costs may exist when the government does not rely on off-the-shelf technology and uses a specific firm with specialized knowledge for software and other purposes.

\textsuperscript{6} Jones and Thompson therefore recommend outlay budgets (e.g., lump sum budgets or input budgets) when confronted by a non-competitive provision of financial resources and unique, hard to measure outputs or outcomes (in Meyers, 1999, P. 578).
on agency behavior. The uppermost line shows the effect of adding additional incentives (e.g., rewards, competitive sourcing). It also may reflect using external production (outsourcing opportunities) and other competition devices. The bottom line shows the consequences of the uses of hierarchy and traditional measures; that is, the more subjective assessments of performance.

When the quality of the performance measurements system is good, it is not difficult to evaluate performance, and consequently, incentivized or market based production is preferable to internal production, since the risk of goal distortion is low. Thus outsourcing using performance contracts or competitive sourcing becomes appropriate when performance measure quality is high.

Figure 1: When competition, outsourcing and block grants produce greater benefits than traditional hierarchy and internal norms

Consider the National Park Service, which some have argued has the largest lawn care management service in the world. This work would be an excellent candidate for competition-
based incentivization or outsourcing. Lawn care contracting is not easily subject to opportunism, the skill and technology levels are low, and competition is high. The federal government cannot easily be extorted. Similarly, to offer technical assistance, the Natural Resources Conservation Service could use local crop consultants and agricultural engineers, rather than rely on the direct provision of services and the hiring of internal staff. The local district conservationists would oversee a region and steer the consultants, rather than directly carry out the technical work himself.\(^7\)

However, when performance measurement quality is weak, adding any type of additional incentives to performance measures leads to poorer agency performance. The effect of the incentives is to encourage greater goal distortion, misdirecting the agency’s efforts and wasting effort. Where program contractability is poor, competition and other incentive devices can steer a program to unintended directions, even if they increase efficiency (Eisenhardt, 1989). With poor contractability, using incentives attached to the performance measures will lead to excessive numbers of Type I and Type II errors. Contracting out when task complexity is high and performance measurement quality is low, is less likely to produce savings or desired results (Globerman and Vining, 1996). As a result, the traditional incremental budget, responding to sectoral needs, coupled with an agency-based personnel system that rewards management skill may be preferable in those circumstances when the quality of the performance measurement system is low.

---

\(^7\) OMB Circular A-76, the 1998 Federal Activities Inventory Reform Act (FAIR), and other government sources have relied on the distinction between activities that are “inherently governmental” and those that are commercial, without being “intimately related to public interest.” A July 29\(^{th}\), 2009 memorandum from Director Orszag to the Departments and Agencies, distinguished between “inherently governmental”, “critical, but not inherently governmental,” and “essential, but not inherently governmental.” Only the former must use only federal employees. However, these distinctions include no bright line differentiations, so in practice, the categorization results in considerable negotiation and some conflict. [http://www.whitehouse.gov/omb/assets/memoranda_fy2009/m-09-26.pdf](http://www.whitehouse.gov/omb/assets/memoranda_fy2009/m-09-26.pdf) [Accessed July 2010]
For example, since education generally produces a poor quality of performance measures, whether the very noisy outcome measures or the very distortive output measures, the measures are less problematic for either information reporting (with improved decisionmaking) or incentivization. As a result, the use of incentives tied to educational standards, and the attempts to motivate school performance through competition from private schools has mostly been ineffective in increasing school quality (McDermott, 2006; Andersen and Serritzlew, 2007). Likewise, No Child Left Behind has been criticized for encouraging schools to ‘teach to the test,’ as well as omitting important curricula that is not covered by the test (e.g., science), and encouraging schools to manipulate the eligibility of the student body in order to increase the score results (Figlio and Getzler, 2002).

In cases where the complexity of the program is great and performance measurement is weak, block or formula grants may be preferable. Instead of directly managing a project, program managers now are overseeing others. The process becomes one of allocating funds to different states according to a formula (potentially with elements that reflect performance-based criteria). Consequently, program managers can effectively deal with programs which have many subprogram elements and goals. They might direct that states allocate funding to the highest ranking projects for each of the various objectives.

On the other hand, the lack of a means to evaluate performance may mean very different results across states. Those states with an intrinsic motivation to see the program succeed may be trusted to use the flexible authority to choose the best way to accomplish the program goals (even if that accomplish is difficult to evaluate). However, states with a less enthusiastic attitude toward the program may see fungible resources as a way to fund programs deemed more
important. Thus the lack of effective performance measures may lead to a failure to accomplish program missions when states receive block granted funding.

The use of poor quality performance measures should be avoided. Even when no incentives are attached to a performance measure, it does not mean that a poor quality measure should even be used as an information source. At best the information is vague; at worst it is distortive and confounds good management efforts.

Sometimes, as a consequence of the complexity and opaqueness of the agency performance measures, delegation with monitoring and incentives may become ineffective as a means of oversight and control. In that case, traditional hierarchical management is preferable for increasing program effectiveness or efficiency.

**Oversight**

Monitoring different aspects of program performance and establishing new incentive devices is costly. Each organizational effort represents either an investment in institutional capacity for monitoring and metering or in regulatory and bureaucratic rigidities. For example, Baker (2000) describes how, because loan officers in a bank are rewarded based on their loan originations, they are motivated to generate riskier loans than if they were rewarded based on the loan’s profitability to the bank. To compensate for use of a distortive performance measure an institutional oversight device is needed to review all the loans proposed by the loan officers. Thus the bank must maintain a separate credit committee to oversee the creditworthiness of the debtors and to approve or deny prospective loans.

In general, the costs of new organizational structures or regulatory changes will rise as we aim for more efficiency. Each additional performance measure imposes additional costs on the government as new organizational monitoring and oversight capacity are necessary.
Different types of programs may benefit to a greater or lesser degree from adding incentives. For some government products, no additional incentive system may be needed or it may not be worth the monitoring cost.

In general, the costs of new organizational structures or regulatory changes will rise as we aim for more efficiency. Each additional performance measure imposes additional costs on the government as new organizational monitoring and oversight capacity are necessary.

Different types of programs may benefit to a greater or lesser degree from adding incentives. For some government products, no additional incentive system may be needed or it may not be worth the monitoring cost.

**Figure 2: Determining Optimal Performance Measurement Systems**

Figure 2 shows that there is an optimal level of performance measures, based on the expected efficiency benefits and oversight or monitoring costs. Efficiency savings from additional performance measures will decrease as the complexity of the cascading incentives
becomes greater and the level of needed to calibrate new offsetting impacts grows. Additional monitoring and oversight is obtained at increasing marginal costs. Direct oversight may partially compensate in the short run for inefficient budget incentives, but with a poor set of performance measures for a program, a lack of structural or institutional incentives will generate inefficient fund allocations and ineffective program administration.

In designing programs, oversight may be built in or waived depending upon the structure chosen. Different budget and program designs require different degrees of oversight. Mandatory programs do not require annual oversight in order to allocate funds. Congressional committees avoid the appropriation process. They avoid the conflicts associated with the decisionmaking process. After an initial authorization, mandatory programs are unhindered in their spending. This permits them to act as entitlements more effectively by ensuring the necessary funds will be available when eligible individuals request it.

Of course, not all mandatory programs are entitlements. Their program design – the tool – for mandatory programs can come in a wide variety of formats: there are mandatory loan programs (US Department of Agriculture’s commodity export loans, PL480, and Department of Education’s Federal family education loan), formula and block grant programs (Department of Transportation’s Highway traffic safety grants, and Emergency preparedness grants, and HHS’ Temporary assistance for needy families, and Social services block grant), and vouchers (USDA’s Food Stamp program).

Similarly, regulatory programs do not require an annual review or reauthorization for their programs. Authorizing committees can establish the terms and conditions for the regulatory regime and then permit the agencies to make decisions on their behalf. The committees avoid routine decision costs, but at the expense of regular and meaningful program oversight. It may be
possible to limit the regulatory program’s authority to a specific term (building in sunset provisions) requiring then a new authorization to proceed. However, in practice the uncertainty generated by the sunset date may undermine the regulation’s goals. In practice the only thing worse, for both the public at large and the regulated community is uncertain or chaotic regulatory requirements.

Methodology

I started with the intersection of two data sets. The Public Budget Database (Budget of the United States Government), which displays all federal programs with the authority to spend (budget authority) from FY 1977-2006. I used the OMB Public Budget database, which includes account-level data for budget authority, 1976-2007, along with account information, such as subfunction code and Budget Enforcement Act (BEA) category of mandatory or discretionary spending.

The programs selected are expenditure accounts, with budget authority available for spending in more than one year. Accounts that were place-holders, intra-governmental transfers, or accounting tools were excluded, along with working capital funds, revolving fund, and liquidating accounts in which funding levels are generated from non-Congressional actions. Accounts with only single year appropriations dropped, since these are usually accounting for intra-government transfers, temporary holding account, or offset, such as blocked mandatory spending. The database includes reorganization effects.

The other data source comes from the Program Assessment Rating Tool (PART). The PART evaluation is based on a questionnaire which rated federal programs. The assessment covered four areas —purpose and design, strategic planning, management, and results and accountability. The answers to questions in each of the four sections produces a 0 to 100 (100 as
the highest) score for each section. The PART assessment also categorizes programs according to the following classifications: Direct Federal, Competitive Grant, Block/Formula Grant, Regulatory, Capital Assets and Service Acquisition, Credit, and Research and Development.

One problem with the PART data is that it does not precisely reflect program accounts, as displayed in the public budget database. In the PART review, the unit of analysis can differ from the formal budget structure of accounts and activities. Instead a “program” may also be a collection of programs or activities that are managed as one entity or that has a clear set of goals. The idea is that when programs share a common goal and are highly interdependent it may make more sense to review them as a unit rather than separately. However, these “programs’ may vary in their program or budget design. That is, a PART program could be composed of both mandatory and discretionary accounts.

For this analysis then, programs used reflect those programs that are common between the two data sets. This significantly narrows the numbers of observations, resulting in 90 observations. There is also a selection effect, but in what way is unclear.

Variables include:

**PART score elements, by section** (see Appendix for additional description):

- S1 score: reflects the degree to which the programs’ design and purpose clear and defensible
- S2 score: involves strategic planning, and weighs whether the agency sets valid annual and long-term goals for programs
- S3 score: rates agency management of programs, including financial oversight and program improvement efforts.
- S4 score: focuses on results that programs can report with accuracy and consistency.
**Delegation/incentivization index.** While some of the classifications by the PART fit into this type of framework (e.g. direct federal), others did not (e.g., research, credit). These latter had to be bifurcated into whether the program was directly applied or delegated. When it came to grants, the degree of fungibility also had to be considered. Thus not only was there the question that the authority to spend was delegated to non-federal sources (high, medium, or low), but how much control the federal agency passed along was also assessed. The index had a scale of 1-5, in which a nearly fully fungible block grant scored a 1, a moderately fungible grant and a regulatory programs a 2, non-fungible grants and delegated credit programs a 3, direct federal provision, including credit and capital asset program, a 4, and competitive grant programs a 5.

**Oversight.** Programs that were either mandatory or regulatory and hence receiving a lower level of required scrutiny were identified.

**Incentivization Index x PART score (S1 or S2):** Interaction term that reflects that the impact of incentivization and delegation as affected by the quality of the performance measurement.

**Dom Invest x DI:** As a control variable: programs were separated into different categories using the budget functional classifications. The four program purposes were: Domestic Investment, International & Defense, Transfers (including health, VA, Income), and Governance. The domestic investment was used as a control variable, interacted by the delegation/incentivization index.

The analysis involved regression analysis using PART scores as the dependent variable. Independent variables included both PART scores, the Oversight variable, the index of incentivization variable, and interaction terms: oversight x incentivization level, PART scores x incentivization index, and domestic incentive program purpose x incentivization.

**Analysis**
In Table I, the program outcome results are regressed on the various elements that affect performance. The important element in this analysis is the impact of the program design or tool on performance. The Delegation/Incentivization Index is significant at $p < .01$, reflecting the importance of the project structure on results. It is negative, indicating that delegation, with fungibility and less incentivization leads to stronger results.

Table I

<table>
<thead>
<tr>
<th>S4 PART score</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Oversight</td>
<td>-98.72* (50.79) -72.13* (38.84)</td>
</tr>
<tr>
<td>Delegation/Incentive Index</td>
<td>-11.8*** (4.24) -14.84*** (2.64)</td>
</tr>
<tr>
<td>Delegation/Incent index x oversight</td>
<td>53.46** (26.29) 40** (20.11)</td>
</tr>
<tr>
<td>Delegation/Incentive Index x S1 score</td>
<td>15*** (.04)</td>
</tr>
<tr>
<td>Delegation/Incentive Index x S2 score</td>
<td>.21*** (.02)</td>
</tr>
<tr>
<td>Delegation/Incentive Index x Domestic Invest</td>
<td>2.39* (1.39) -1.59 (1.06)</td>
</tr>
<tr>
<td>constant</td>
<td>47.17*** (10.09) 48.57*** (7.69)</td>
</tr>
<tr>
<td>Obs</td>
<td>90</td>
</tr>
<tr>
<td>R_squared</td>
<td>0.17  0.52</td>
</tr>
</tbody>
</table>

*** $p < .01$, ** $p < .05$, * $p < .1$

This result is tempered based on the nature of the oversight regime and of the performance measurement system. When the variable associated with performance measure quality (PART scores) is interacted with the delegation/incentivization, the result is significant at $p < .01$. This indicates that the impact of incentivization depends on the quality of the

---

8 The difference from including both PART section (S1 and S2) variables together in the regression is insignificant:

<table>
<thead>
<tr>
<th>s4_score</th>
<th>Coef. Std. Err. t P&gt;</th>
<th>t</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oversight</td>
<td>-72.82316 38.98978 -1.87 0.065 -150.3723 4.725944</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DI index x Oversight</td>
<td>40.16957 20.17902 1.99 0.050 .0343119 80.30483</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DI index x S2</td>
<td>.1980664 .0254422 7.78 0.000 .1474629 .2486699</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DI index</td>
<td>-16.11237 3.288743 -4.90 0.000 -22.65355 -9.571197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dom Invest Program x DI</td>
<td>-1.659327 1.06441 -1.56 0.123 -3.776396 .577426</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DI Index x S1</td>
<td>.0223391 .0341318 0.65 0.515 -.0455477 .0902258</td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>48.63816 7.720103 6.30 0.000 33.28319 63.99314</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj R-squared</td>
<td>0.5133</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---
performance measurement system, as predicted. This was tested with both the PART’s S1 (design) and S2 (strategic planning) scores to assess different aspects of performance measurement. Only when the performance measurement system is satisfactory will incentivization be effective. They can be detrimental should performance measures be very weak. We also see the difference between the two measures on program outcomes. The adjusted R_squared is 17% for the PART section S1, while the PART section S2, which emphasized strategic planning and measurement design, was 52%.

We also see that the impact of budget design – the case of when the delegation index is interacted with oversight -- to be significant at p < .05. Thus, the need for and impacts of oversight will differ depending on the tool or program design. We see that the impact of incentivization or delegation on program results is affected by the nature of the oversight. The decision to select a particular program type must be tempered by the expectations over the expected oversight regime.

We also see that the level of oversight alone is negative and significant at 1% indicating that more routine oversight leads to better performance. The type of program had an influence on outcomes when interacted with the nature of the program tool. The resulting interaction term was slightly significant at p < .1, with delegation leading to better results for the domestic investment programs.

We can see the relative impacts of the different variables when we standardize the coefficients (see Table 2). The predicted mean score decline in the desired outcomes of .63 comes when we have a standard deviation’s increase in delegation/incentivization. In general, greater incentivization or more controlled delegation reduces desired outcomes.
On the other hand, that result changes when incentivization is conditioned by performance measurement. The DI Index x S2 score interaction has the highest impact on outcomes. Each one standard deviation increase in the interaction of delegation/incentivization as interpreted through performance measurement quality leads to a .97 standard deviation increase in program results. Likewise when the degree of delegation is interacted by oversight, for each 1 standard deviation change in this variable we see a .81 standard deviation increase in program results.

Table 2

<table>
<thead>
<tr>
<th>S4 PART score</th>
<th>Coefficient (se)</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oversight</td>
<td>-72.13* (38.84)</td>
<td>-.77</td>
</tr>
<tr>
<td>Delegation/Incent Index</td>
<td>-14.8*** (2.64)</td>
<td>-.63</td>
</tr>
<tr>
<td>Delegation/Incent index x oversight</td>
<td>40** (20.11)</td>
<td>.81</td>
</tr>
<tr>
<td>Delegation/Incent Index x S2 score</td>
<td>.21*** (.02)</td>
<td>.97</td>
</tr>
<tr>
<td>Delegate/Incent Index x Domestic Invest</td>
<td>-1.59 (1.06)</td>
<td>-.11</td>
</tr>
</tbody>
</table>

*** p < .01, ** p < .05, * p < .1

Block grants are often lauded for involving local jurisdictions and getting closer to the ground. However, for the approach to be meaningful, local authorities seek the freedom to respond to local needs and not be hamstrung with federal controls and motivational devices that steer them one way or another. The principle of subsidiary suggests that funding granted to local authorities should be more fungible and not be tied down with federal preferences, but that we should enable the local authorities to react according to immediate opportunities that present themselves. However, we can see that fungibility has its limits when performance measure quality is good. In those circumstances, competitive grants and more federal controls are preferable.

The results are similar for outsourcing, competitive sourcing, and related incentivized approaches. Outsourcing, when competitively awarding contracts is likely to lead to improved
results, in those cases where the performance measurement regime is excellent. It is not surprising therefore that we would see excellent results for many of the New Public Administration results when carried out at the local level with simple tasks, which are easily evaluated, including relying on private sector or other local municipalities as benchmarks (Brown and Potoski, 2003).

It may also reflect the prevalence of outsourcing for activities within an agency. For example, the Forest Service engages in many activities for which performance measures are confusing or weak. It is difficult to evaluate forest health or ecosystem conditions, since sometimes the gain for one habitat or species is the loss for another. However, other tasks or functions are routine without any judgment, so that performance measurement quality can be high. Preparing a NEPA document that captures different potential types of land treatments is an activity that many private consultants can do. This work, like road construction or food and drink concessions, is highly contractable and so amenable to outsourcing.9

Direct provision by federal employees should not be dismissed as archaic or passé. Federal employees may reflect a selection outcome –They are more motivated by the public good outcome than non-federal employees. Their public service interest is part of what drove them to their jobs. Thus direct provision offers the advantage of personnel who have selected into that job in order to carry out program purposes.10

This discussion of contracting out is considering it when it is done competitively. Outsourcing to a non-competitive market would not be expected to yield greater rewards,

---

9 A risk with outsourcing also comes from the loss of memory and learning opportunities for agency. The more specialized knowledge builds up, the more of a specific asset is created with the risk of opportunism and hold-up costs. These transaction costs may be a motivation for not out-sourcing the product. The less specific knowledge and skill developed, the more generic the product, and the larger the competitive market, the greater the scope for outsourcing.

10 Of course their zeal for the job may be reflected in also pursuing increased appropriations in order to accomplish the agency missions.
regardless of the performance measures. According to a March 4, 2009 Memorandum from the President, “Since 2001, spending on Government contracts has more than doubled, reaching over $500 billion in 2008. During this same period, there has been a significant increase in the dollars awarded without full and open competition and an increase in the dollars obligated through cost-reimbursement contracts. Between fiscal years 2000 and 2008, for example, dollars obligated under cost-reimbursement contracts nearly doubled, from $71 billion in 2000 to $135 billion in 2008.”

Agencies may be resorting to the use of cost-reimbursement contracts where program contractability is low, thereby avoiding the risks from incentives. However, the inability to monitor and meter the private contractor due to asymmetric information suggests that government agencies are likely to overpay for these contracts and would do better relying on internal production. In general, sole source contracts yield all the worst possibilities of both direct provision and outsourcing. In addition, there is additional risk of corruption or federal employees ingratiating themselves for their next job.

Conclusion

Think of the Northern Cardinal. They are a bright red bird, so bright as to be visible across a large woodlot in the winter. Certainly, it does make the male Cardinal quite visible to other cardinals, both male and female. That helps for attracting females and guarding territory. However, it does not seem to be an ideal way of avoiding predators. One may ask whether this is the optimal Cardinal design. Would a more drab cardinal succeed just as well? Perhaps, blue, striped or iridescent Cardinals would be more effective in practice. Is the design of the Cardinal optimal? Would other designs serve the Cardinal equally as well, with red just being one, among

---

many possible equilibrium outcomes? In fact, maybe the design of the cardinal isn’t even an optimal design, but rather a less efficient, but stable alternative.

We can ask the same types of questions as to the design of governmental organizations. Too often choices in administrative design or budget systems seem to the product of fads, not evidence. States or national governments would adopt new administrative design or budget systems even though evidence of results was always hard to come by (since measuring results was itself a challenge). In reality, just agreeing on the preferred outcomes for a program had never frequently been done. So, administrative tools have been adopted in helter skelter fashion, in some states and federal departments wholeheartedly; while in others, haltingly and reluctantly.

This paper attempts to move beyond the previous position. There is a saying that “To the man who only has a hammer in the toolkit, every problem looks like a nail.” (Abraham Maslow). The federal government does not have only a hammer. It has a full tool chest. The problem has been in knowing which tool is appropriate for the occasion. This paper will hopefully help with that judgment. By assessing the quality of the performance measurement regime, one can then choose the preferred type of administrative design.
Appendix

Program Assessment Rating Tool (PART) Elements by Section

Section I: Program Purpose and Design

- Clarity and relevance of program purpose
- Soundness of program design
- Addresses program’s structural issues
- Clear design and purpose an essential for identifying performance measures

Section II: Strategic Planning

- Addresses program’s plans and approach to achieve specific long-term goals
- Programs must have long-term and annual performance measures
- Programs must have ambitious targets
- Evaluation of program effectiveness and to support performance improvement

Section III: Program Management

- Addresses elements related to managing a program to achieve performance goals
- Accountability of managers, performance of partners
- Coordination with related programs
- Financial management, improving efficiency
- Addressing deficiencies

Section IV: Program Results/Accountability

- Assesses the extent to which a program is achieving its long-term and annual performance goals and efficiency goals
- Reporting of actual performance compared to targets (identified in Sections II & III)
- Effectiveness in achieving goals based on independent evaluations
- Comparison of performance with similar programs
WORKS CITED


