Political Appointees and the Quality of Federal Program Management*

In this paper we use the Bush Administration's management grades to analyze whether programs administered by senior executives (SES) are better managed than those administered by political appointees requiring Senate confirmation (PAS). We explain the administration's management grading scheme and how it can be informative for evaluating comparative management quality. We explain why senior executive-run programs should be better managed than appointee-run programs and test our claim with data on 234 federal programs. We find that political appointee-run programs earn systematically lower grades in most management areas. We conclude that a systematic reconsideration of the proper role of political appointees in federal program management should be considered.

John B. Gilmour
College of William & Mary
Department of Government
P.O. Box 8795
Williamsburg, VA 23187-8795
jbgilm@wm.edu
(757) 221-3085

David E. Lewis
Princeton University
Woodrow Wilson School of Public and International Affairs
311 Robertson Hall
Princeton, NJ 08540
delewis@princeton.edu
(609) 258-0089

---

* We thank Steve Balla, Greg Huber, and Nolan McCarty for helpful comments. We thank the Office of Management and Budget for making their management grades publicly available. The errors that remain are our own.

In 2003 the highly publicized Volcker Commission report recommended among other things that the executive branch be reformed to reduce the number of political appointees.[1] It also recommended that operating agencies be run by managers chosen for their skills and expertise. One of the concerns of the commission was that federal programs and agencies run by political appointees suffer from poor management. Political appointees often do not have the management skills, policy expertise, agency experience or working networks that facilitate public management. Federal programs fortunate enough to be managed by political appointees adept at public management still experience regular turnover at the top.[2] This turnover can have a number of detrimental effects both inside and outside the agency. Frequent turnover creates leadership vacuums, a lack of continuity, and mixed signals about agency and program priorities. Turnover also disrupts working relationships among functionally related agencies and programs achieved through inter and intra-agency working groups.

The Volcker Commission is not alone in noting the potential harms the large number of political appointees can have on management. The recent legislation creating the Department of Homeland Security follows the advice of a 2002 Brookings Institution report and includes explicit limits on the number of political appointees. Similar limitations were included in the 1989 legislation creating the Department of Veterans

---

[1] National Commission on the Public Service. 2003. *Urgent Business for America: Revitalizing the Federal Government for the 21st Century*. Washington, DC: Brookings.

[2] See Brauer (1987), Chang, et al. (2000, 2003), General Accounting Office (1994), Mackenzie (1987), Stanley, Mann, and Doig (1967).

Affairs (Light 1992).[3] This year Congress has held hearings on the possibility of reducing the number of political appointees and last session Congressman Bill Luther (D-MN) introduced legislation to cap the number of political appointees at 2,000. There are currently approximately 500-600 appointees requiring Senate confirmation (PAS positions).[4] According to the Office of Management and Budget, political appointees administer some federal programs and senior executives, either career or non-career, run the vast majority of the rest.[5]

Despite claims that programs run by political appointees are worse off than programs run by managers drawn from the Senior Executive Service, there is no consensus that this is the case. Presidents historically have jealously guarded their ability to make appointments in the executive branch. Attempts to elevate the Environmental

---

[3] See Daalder, et al. (2002).

[4] The Brookings Institution's Presidential Appointee Initiative identified 501 positions but this is a bit low compared to past estimates. A 1994 General Accounting Office report identified 567 positions. The 2000 Plum Book identifies 1,203 PAS positions but includes minor appointments to various small commissions, ambassadorships, and Justice Department positions that are Senate-confirmed by tradition, like U.S. Marshals and U.S. Attorneys.

[5] According to the Office of Management and Budget, in the sample of federal programs that is the subject of this study, political appointees administer 165, or 77%, of the programs. Career SES managers direct 20 (9 %), non-career managers run 28 (13%), and 1 program was administered by someone excepted from the competitive service by statute (Student Aid Administration).

Protection Agency to a cabinet department in the presidency of George H.W. Bush foundered partly on disagreements about the number and depth of political appointees in the proposed department.[6] As the administrative state has grown and delegation of policy making authority to the executive branch has increased, presidents have sought to shape policy outcomes administratively through an appointments strategy (Nathan 1975, 1983; Moe 1982, 1985). The number of management positions has increased over time, many of them politically appointed (Light 1995). While no work to date has connected the increasing management positions directly to presidential action, it is reasonable to conjecture that presidents prefer more control over the administrative state to less.

For administration officials an important component of good management is responsiveness to the policy program of the chief executive. That is, a public manager cannot be considered a good manager if they do not take the policy prescriptions of the administration and translate them into concrete administrative goals to be implemented. While ideally career managers would follow the ethos of "speak up, shut up, carry up, carry out," presidents are dubious that they can do so.

So, are political appointees good managers? Are they better managers than Senior Executives? The answer to this question would involve some combination of their ability to take political direction, translate it into a clear program, and implement it consistently and effectively over time. Answering this question is complicated by the fact that it is extremely difficulty to measure comparative management quality. Agencies have multiple and conflicting goals that are difficult to measure in a systematic way. When multiple agencies are considered side-by-side the complications increase since

[6] See *Congress and the Nation, v. 8, 1989-1992.* Washington, DC: CQ Press. p. 498.

different agencies have unique mandates and responsibilities. Doing a comparative study of management quality is a huge undertaking. It requires, among other things, a definition of what good management is and what programs will be analyzed. It requires that these determinations are sensitive to differences across agencies and it requires retrieving information from agencies that may not have an interest in providing the information.

It is in this context that the Bush Administration's recent attempts to grade the management of federal programs provides a unique opportunity for scholars to study federal management. The Bush Administration numerically graded the management of 20 % of federal programs, or 234 programs, for the FY 2004 Budget and made these grades publicly available in budget documents and on their website. These 234 programs represent $480 Billion, or 24% of federal spending.[7] The stated intention of the administration is to use these management grades to determine increases or decreases in each program's budget. The grades, referred to as PART scores, come on the heels of a less ambitious and less detailed attempt to grade 100 programs for the FY 2003 budget.[8]

In this paper we use these grades to compare the management of federal programs by political appointees and managers from the Senior Executive Service. There are numerous and obvious objections to using these grades for this purpose and we deal with these concerns in the first section of the paper. In the second section we explain why

---

[7] U.S. Office of Management and Budget. *M-02-10. Memorandum for Heads of Executive Departments and Agencies*. Washington, DC, July 16, 2002.

[8] For a listing of programs and program grades for both FY 2003 and FY 2004 go to http://www.omb.gov.

programs run most proximately by career executives are likely to get better management grades than programs run by appointees. In the third section of the paper we describe the grades in more detail and analyze them to determine whether there are significant differences in the management grades of political appointees and senior executives. We find that programs run by political appointees do significantly worse in all but one category of management grades and in that one there is no significant difference between programs run by political appointees and programs run by executives from the SES. Importantly, we find that programs run by non-career appointees to the SES are managed comparably to programs run by career employees in the SES. In the fourth section we discuss the results, their implications, and our conclusions.

I. Can PART Scores Tell Us Something Reliable About Management Quality?

The Bush Administration's management grading scheme is relatively straightforward. It was designed by OMB in consultation with the President's Management Council, an advisory council of lower level agency political appointees, and includes numerical grades from 0 to 100 in 4 categories and a final total weighted numerical management grade. The four categories with their purposes are:[9]

[9] U.S. Office of Management and Budget. *Instructions for the Program Assessment Ratings Tool*. Washington, DC, July 12, 2002. See also U.S. Office of Management and Budget. 2003. *Budget of the United States Government FY 2004: Performance Management and Assessments*. Washington, DC: U.S. Government Printing Office.

1. Program Purpose & Design (weight= 20%): to assess whether the program design and purpose are clear and defensible

2. Strategic Planning (weight= 10%): to assess whether the agency sets valid annual and long-term goals for the program

3. Program Management (weight=20%): to rate agency management of the program, including financial oversight and program improvement efforts

4. Program Results (weight=50%): to rate program performance on goals reviewed in the strategic planning section and through other evaluations

Grades were determined in each category based upon answers to a series of yes/no questions relevant to the section in question and adjusted for the type of program under consideration (block grant, regulatory, credit, etc.). For example, one question used to assess the quality of strategic planning asks, "Does the program have a limited number of specific, ambitious long-term performance goals that focus on outcomes and meaningfully reflect the purpose of the program?" For this and other questions the OMB provided background information on the purpose of the question and elements of an affirmative response. Answers were determined jointly by the agency running the program and an OMB examiner. In cases of disagreement they were resolved through arbitration by OMB hierarchy, namely the OMB branch chief and, if necessary, the division director and Program Associate Director.

Some Possible Objections

Performance measures of any type are controversial in public management. There are a number of potential objections to using them to make any conclusions about public management at all. First, the grades could be politicized. The administration could assign grades to programs on the basis of ideology or partisanship rather than on true management merit. It is possible, for example that programs that are typically "Democratic" programs like environmental, regulatory, or social welfare programs are graded systematically lower than other programs. Second, the grades could be inaccurate, assigned on the basis of a subjective opinion rather than true management quality. Graders could use the wrong categories for determining what good management is. They could also have the wrong people do the grading. OMB examiners unfamiliar enough with particular programs could make big mistakes in grading.

All of these are serious concerns but <u>even if these concerns are true</u>, it does not eliminate the usefulness of the grades for making inferences about the comparable management quality of programs run by appointees or senior executives unless certain other conditions are true. Let's start with a general question about the grades. If one program were objectively horribly managed and another was managed extremely well, would this be picked up in the grades on average? If one program's purpose or mandate was really murky and one program's purpose was crystal clear, would this be reflected in the numerical grade under program purpose and design? If one program used principles of strategic management to guide implementation and another did not, would this be picked up on average? The answer to those simple questions is probably yes even if there was error, bias, or favoritism.

In statistical terms, this means that true management quality is a function of factors that add up to true management quality and some error, either a little or a lot. This error is not a problem for making comparisons about management quality between different types of managers unless the error is non-random and not only non-random, but also correlated with whether or not a program is run by a political appointee or senior executive. If the error in management grades is purely random, say a few extra points here and a few less there for all the grades, there are statistical techniques that can distill out the effects of factors we care about from the error. Students of statistics can picture a regression line surrounded by normally distributed points that reflect observed values. What if, however, the errors are not random? What if programs created by Democratic presidents are graded systematically worse than programs created by Republican presidents? While not an admirable grading scheme, this would not cause problems for our inferences so long as Democratic and Republican programs were equally likely to be run by political appointees or senior executives. That is, if the errors in grading are uncorrelated with the variable of interest (political appointee vs. senior executive) we are on solid ground using the administration's management grades to make inferences about the quality of management. We can also estimate statistical models that allow us to control for other hypothesized determinants of management grades including ideological content, program type (e.g., block grant vs. regulatory), etc.

What if the programs themselves were selected ahead of time to make a point? Suppose, for example, that only politically-headed programs that were well run were chosen? This type of sample selection problem would be a serious issue if the administration made choices about programs this way. We have no evidence, however,

that there was bias in the selection of federal programs. The administration claims loosely that a stratified sampling scheme was used. That is, they looked for variance in program selection. In particular they looked for both large and small programs, programs with a history of good management and bad management, and programs of varying types and purposes. They also chose for inclusion programs that were recommended for review by interested parties and programs coming up for reauthorization.[10] The administration is adding another 20% of federal programs into the grading scheme for the FY 2005 budget, and suggesting that their intent is to grade all federal programs.

It is our view that these management grades provide a unique opportunity to examine important questions in public management if properly utilized. Based upon our analysis of these grades, there appears to be no reason why they cannot be used to evaluate comparable management quality among programs even if there is error or bias in the grading.

II. Federal Programs, Management Status, and Quality

Some program management problems arise because levels in the management chain have different priorities or preferences. These differences in priorities or opinion arise commonly in two ways. They may be due to inherent ideological differences among the levels. For example, program employees may not share the ideological beliefs of the manager and therefore implement the manager's goals with less enthusiasm.

---

[10] As described by Marcus Peacock, Program Associate Director, Office of Management and Budget in a forum entitled *Program Performance and the FY 2004 Budget Process*, June 13, 2003, 2247 Rayburn House Office Building, Washington, DC.

Preference divergence can also arise because of unclear communication among the levels. Program managers may not know what politicians want, particularly when statutory language is unclear or politicians within or between branches disagree themselves. Similarly, program employees may deviate from the direction a program should take simply because they are not clear about what they should be doing or how to change what they do to align their actions with program goals.

Of course, some of the problems with differences of opinion or disagreement among the levels can be resolved through monitoring. For instance, if managers have a good working knowledge of an agency or can easily observe and understand what program employees are doing, they can induce program employees to do what the program requires. There is a robust literature on the political oversight of bureaucratic actors that describes the importance of monitoring in inducing compliance. Monitoring is equally important in the second stage of the management chain.

Slippage can also occur in the politician-manager-program chain because one of the links in the chain is weak. That is, the wrong people will have been selected or chosen for a particular position. Of course, the weak link may be the first one. Some programs may be poorly designed in law or statute, damaging prospects for success from the beginning. It may also be the last one. Programs may lack capacity to do what they are supposed to do. They may be understaffed, underfunded, or under-equipped. Holding these possibilities constant, however, there is also variation among programs in the strength of the management part of the chain. Some managers may perfectly understand what they should be doing, perfectly willing to do so but lack the capacity. For example, politically appointed managers may know what the administration wants or

even what all the stakeholders want. Yet, they may lack the ability to translate that information into a clear program purpose, use strategic planning to design a program, manage the program, and measure results to demonstrate the success or failure of the program.

In important ways senior executive-run programs as a class have systematic advantages over political appointee-run programs in all three of these areas. They are less likely to suffer from different priorities and goals at the different levels. Senior executive-run programs are better able to translate the interests of stakeholders into a clear, consistent program purpose. While political appointees are more likely to have relationships with the administration, administration appointees, and groups outside Congress and the administration that share the administration's views, they are less likely as a group to understand and have relationships with the whole panopoly of stakeholders whose support is necessary for craft a stable program purpose. Senior executives are more likely to serve through multiple administrations, develop relationships with key officials in OMB, congressional committees, and interest groups, and have the skills to broker an acceptable definition of program purpose.

Senior executive management continuity helps programs craft and communicate clear goals to program employees over a longer period of time. Frequent turnover among political appointees, however, creates leadership vacuums in federal programs. Inexperienced newcomers or acting assistants with little leverage are more likely to run these programs. This creates discontinuity in program goals, ineffective oversight, and malaise among career employees (see, e.g. Ban and Ingraham 1990, Heclo 1977, Joyce

1990, Mackenzie 1987). Recent criticisms of Department of Energy security lapses, for example, name frequent appointee turnover among the primary causes.[11]

Senior executives can also generally do a better job overseeing their programs because they are more likely to be drawn from that program or a related one and have substantive expertise in the program area they manage. They are more likely to understand the program and its requirements and understand what change is possible and the ways to achieve it. Because of their experience and personal relationships inside the program, they are better able to oversee the program, its processes, and personnel and bring the perspectives of employees in line with larger program goals.

Senior executives are also more likely to have public management experience. One of the frequent criticisms of political appointees is that they have never been trained in public management or had any experience managing a large public organization prior to their service. Senior executives, on the other hand, have the advantage of public organization experience, training, and a knowledge of public organizations and how they work. In order for federal programs to score well on the PART grading system, managers must understand how to set clear goals, design programs (including financial management and e-government), develop adequate performance measures, and use these measures to demonstrate results. Public managers drawn from the SES are more likely to have these skills, being trained as public managers formally or learning through experience on the job.

---

[11] President's Foreign Intelligence Advisory Board. 1999. *Science at its Best, Security at its Worst: A Report on Security Problems at the U.S. Department of Energy*. June, 1999.

Of course, some programs are extremely fortunate and are administered by political appointees of extraordinary ability. These appointees are able to craft a clear program purpose and engage in strategic planning, program management, and performance measurement by virtue of a longer than average tenure, good program knowledge, and good management skills. There are fewer programs, however, that benefit from multiple political appointees successively of just this type. Managers from the Senior Executive Service are more likely to have these skills and programs run by senior executives are more likely to have managers of this type in succession.

III. Data, Variables, and Methods

Our expectation is that senior executive-run programs run should get systematically higher grades than appointee-run programs in all four management categories—program purpose, strategic planning, program management, and program results. This should also be evidenced in the total weighted score and overall categorical classification as ineffective, results not demonstrated, adequate, moderately effective, and effective. We examine the 234 federal programs assigned PART scores for the FY 2004 budget to test our expectations.

The Grades

In Figure 1 we graph the distribution of categorical grades. Over half were graded either "ineffective" or "results not demonstrated." The categorical grade and the total numerical grade were determined by raw scores in 4 areas that were weighted and summed. We graph the average scores in all 4 areas along with the total score in Figure

2. The lowest scoring program overall received an 11 and the maximum was a 93 out of 100. The average scores was 59.76 (sd 16.78). The disability compensation program in the Department of Veterans Affairs (11/100) and the IDEA Preschool Grants Program in the Department of Education (19/100) were the two lowest scoring programs. Broken down by category, the average raw score for program purpose and design was 84.9 (sd 18.36) with a minimum of 20 and a maximum of 100. The mean score for strategic planning was 67.09 (sd 22.16) with a minimum score of 0 and a maximum score of 100. For program management the mean score was 71.7 (sd 18.95) with a minimum and maximum of 10 and 100. The program results average was 43.7 (sd 24.3).

Insert Figure 1 here.

By far the highest average grade was given for program purpose and design which suggests that this grade might be less informative than the other grades. Indeed of the 231 programs for which we have a grade in this category, 112 received a grade of 100. Not surprisingly, program results had the lowest average score (43.7) since this score is partly determined by success in the other 3 categories.

Insert Figure 2 here.

PAS or SES

To determine whether a program was administered by a political appointee or senior executive we consulted OMB worksheets which list the bureau administering each program.[12] We then looked up the bureau in what is commonly referred to as the 2000

[12] Source: U.S. Office of Management and Budget. 2003. *Budget of the United States Government FY 2004: Performance Management and Assessments*. Washington, DC: U.S. Government Printing Office.

"Plum Book" to determine the appointment status of the bureau chief.[13] One complication with this approach is that the career/non-career status of SES bureau chiefs can change.[14] Many programs will have been run by both career and non-career SES managers at different points in their history. This makes it difficult to distinguish career SES-run programs from non-career SES-run programs and consequently difficult to asses the relative management quality of career and non-career managers. We report data on non-career SES versus career SES and political appointees with this caveat.

Of 214 programs for which data was available, 165, or 77%, were administered by bureaus headed by political appointees. Of the remaining 49 programs, career employees of the Senior Executive Service administered 20 and non-career SES employees or officials excepted from competitive service requirements by statute directed 29. The programs administered by bureaus headed by political appointees had significantly lower management scores than other programs. Figure 3 includes a graph of average program scores by appointment status of the bureau chief. There are no

---

[13] Source: U.S. Congress. Senate. Committee on Governmental Affairs. *Policy and Supporting Positions.* 106th Congress, 2nd Sess.

[14] Of course, in a few cases whether a position requires Senate confirmation or is staffed by someone drawn from the SES is discretionary. The number of such cases is limited to those agencies where statutes prescribe limits on the number of Senate-confirmed appointees but do not say which positions they are. This is a relatively new statutory innovation and those positions that have changed usually involve process bureaus like legislative affairs, administration, or procurement rather than policy.

significant differences in the clarity of program purpose or design (the first bar) although programs run by career SES employees have a slightly higher score on average. In the other 3 categories and the total score, however, SES officials have significantly higher scores than presidential appointees. The differences between the scores of PAS appointees and other agency heads is between 5-8 points and statistically significant in difference of means tests ($p<0.05$).

Insert Figure 3 here.

On its face this supports the conclusions of the Volcker Commission and others that programs run by political appointees often suffer either because of a lack of management skill or because of frequent turnover. This is particularly interesting because if political bias exists in the grades, it should work in the other direction. The administration should seek to grade their own political appointees higher than career employees. The results, however, suggest that even with this potential bias, SES-run agencies get higher grades.

It is worth noting that while there is a significant difference between political appointments that require Senate confirmation and SES managers, there is virtually no difference between career and non-career SES managers in the scores. This may reflect the fact that non-career SES appointees are drawn from a different population of potential managers than higher level political appointees. They may have, on average, more previous federal experience, more managerial or substantive expertise, or a less politicized view of their job as program managers than their political counterparts. It is also possible that the indistinguishability of career vs. non-career SES reflects that fact

that career status of SES program managers changes depending upon the choice of the current administration.

Regression Analysis

There are a number of possible objections to the relationship described above between SES/PAS management and management grades. It is possible, for example, that certain types of programs are easier to govern than other types of programs and that the type of program is also correlated with the type of manager. It is possible, for example, that block or formula grants are easier to administer than regulatory programs and grant programs tend to be run by career managers rather than political appointees.

Insert Figure 4 here.

Another possible objection is that grades differ by department and the level of political appointee penetration also varies by department, thus giving the appearance of a relationship where there really is not one. This clustering of programs in departments might lead us to falsely attribute to the difference between SES/PAS what is actually caused by factors unique to a department. It is possible that poor management performance could be a function of the larger department, its culture, its history, and the difficulty of its task than anything intrinsic to the difference between SES managers and PAS managers.

Insert Figure 5 here.

To address these possibilities we turn to regression analysis that will allow us to control for some of the possible confounding factors. To account for the differences in program type, we include dummy variables for each type—competitive grant,

block/formula grant, regulatory, capital assets and service acquisition, credit, direct federal, and research and development programs.[15] We also include indicators for each agency that houses a program.

Some of the difficulty or ease of managing a program may be due to the politics at the time the agency was created and the length of time the program has been in existence. Moe (1989) argues that the design of federal programs is the result of a struggle among legislators, interest groups, and presidents. The opponents of new proposed federal programs do not want them to succeed and, to the extent they have influence, add features that will ensure their failure. The difficulty of managing federal programs, then, may be a function of the level of disagreement at the time the program was created. To test for this we include an indicator for divided government at the time of the program's creation. Our expectation is that divided government produces programs that are more difficult to manage because of conflicting prescriptions, vague legislation, and unclear goals embedded in law.

To account for the politicization that may be present in grading, we include indicators for the presence of a Democratic president (0,1) and a Democratic Congress (House and Senate, 0,1). If the grading is really politicized, programs created under Democratic presidents and Democratic congresses should receive lower grades than other programs.

Sometimes program performance can be a function of other characteristics of management that have little to do with whether the program is run by someone drawn

---

[15] Source: U.S. Office of Management and Budget. *Instructions for the Program Assessment Ratings Tool*. Washington, DC, July 12, 2002.

from the SES or an appointee chosen by the White House. Three features of agency design are relevant here. First, sometimes when politicians are concerned about management they give agency officials fixed terms to ensure both long tenure and depoliticized management.[16] To account for this we include an indicator for whether the bureau chief in question has a fixed term.[17] Eighteen of the evaluated programs were run by agencies whose heads served for fixed terms. Second, the size of a federal program may have some impact on the ease of management. To account for this we have estimated models that include the natural log of the program budget.[18] Budgets vary in size from virtually no budget to upwards of $30 Billion (Highway Infrastructure). We do not include estimates of models with the budget data in the main text because of missing data but include the estimates in Appendix A. Finally, some programs are much older than other programs.[19] The patent program and the trademark program, for example,

---

[16] One recent example is the creation of the National Nuclear Security Administration in 1999. Congress granted the NNSA administrator a fixed term to ensure longevity. See Lewis (2003).

[17] Source: U.S. Congress. Senate. Committee on Governmental Affairs. *Policy and Supporting Positions.* 106th Congress, 2nd Sess.

[18] Source: U.S. Office of Management and Budget. 2003. *Budget of the United States Government FY 2004: Performance Management and Assessments*. Washington, DC: U.S. Government Printing Office.

[19] We did research on each individual program to determine the date of origin. Many of the program dates could be obtained by finding their original authorizing statute. Sometimes origin dates were included in the PART worksheets. In other cases, we

trace their history back at least 200 years. Our expectation is that older programs should be managed better due to experience, institutional memory, and learning over time.

Methods

We estimate models both of the numerical PART scores and the overall categorical grade. One difficulty with the data is that the raw scores have unique distributional properties. In particular, the raw scores often cluster around the maximum or the minimum (lots of 0 scores or 100 scores). This can lead to heteroscedastic errors and biased estimates. One way to resolve this problem is to take the log odds of the score and toss out the scores at the extremes. This will solve the clustering problem but will eliminate information unnecessarily. Another way of resolving this problem, and the one we choose here, is to estimate tobit models where necessary to take into account the limits on the distribution of the data.[20] In doing so we are implicitly assuming that those scores clustering on the endpoints might be distributed further out on a real scale if we could observe them. For the categorical grades we report estimates of an ordered probit

---

researched programs on the web and in government documents including the *United States Government Manual* and the *Federal Register*. We also consulted publications such as the CQ *Federal Regulator Directory* and other guides to federal programs.

[20] The distribution of the total weighted grade appears more or less normally distributed with no clustering on the end points.

model of outcomes.[21] We note, however, that the type of model makes little substantive difference for the estimates.

IV. Results

We include our model estimates in Table 1. In general, the models perform well and confirm what our bivariate analysis suggested, programs run by non-SES appointees are graded significantly worse on different dimensions of public management. The table includes 6 columns, 1 for each management component, a column for the total numerical score, and final column modeling the categorical grade.

Insert Table 1 here.

The one model that does not perform particularly well is the model of program purpose. None of the hypothesized variables was significant. Only the fixed effects for program type were significantly related to the score. Block grant and competitive grant programs (along with acquisition programs) are systematically graded lower on program purpose than other programs. It is surprising that the creation of a program in divided government does not lead to a lower score on program purpose and design. There are several possible explanations. First, it is possible that grading on program purpose is just not very informative. Indeed, there are a lot of "100" grades given and the mean score on program purpose is higher than on other grades by more than 10 points. Second, it could be that OMB examiners and agency officials take into account statutory vagueness when assigning grades. That is, they recognize implicitly that some programs have multiple

[21] We have also estimated multinomial logit models to verify the robustness of the ordered probit findings and the results do not change.

and conflicting purposes and give relative, rather than absolute grades. Third, it could be that this is the easiest part of the federal grading system for managers to manipulate. Since the Government Performance and Results Act, managers have had lots of experience writing successful statements of mission and purpose. Finally, it is possible that the importance of divided government for predicting clarity of program purpose diminishes over time. We have done some additional analyses that suggest this might be the case but the results are not very robust.[22]

After taking into account program purpose and design the estimates confirm much what we expected. In particular, the coefficient on PAS appointee is significant and negative in the 5 remaining models indicating that programs administered by political appointees get lower grades. If a program is administered by a bureau with a PAS appointee in charge, the raw scores on strategic planning, program management, and program results are estimated to be 10-15 points lower, quite a substantial amount. The total weighted management grade is close to 10 points lower for the PAS-run programs. Programs run by political appointees are 35 % more likely to be graded as ineffective or as not having demonstrated results. They are 6 % less likely to be graded as effective.

This is an important finding. Our results cannot arbitrate among competing explanations for why political appointees might get worse management grades. There are several possible reasons that are not mutually exclusive. Appointed managers often have less managerial experience, less substantive expertise in the area they are managing, and

[22] In particular, we have estimated models that include an interaction of duration and divided government. The coefficients are significant or close to significant in one-tailed tests in some specifications but not in others.

less on the job training than their counterparts in the SES. Appointee-run programs suffer from frequent turnover at the top and this turnover leads to an accumulation of management problems. Over time these programs become increasingly difficult to manage. A recent study by Chang et al. (2003) finds that the average appointee serves for a little over 2-3 years when the average CEO serves from 5-7 years.[23] This is the first systematic evidence we are aware of that demonstrates the adverse consequences of political appointee management.

It is important to note that these models could actually underestimate the true impact of PAS administration if bias does exist in the assignment of management grades. To illustrate this possibility, imagine that the administration artificially inflated the management grades of PAS-run programs by 5 points. Since the effect is linear, we would then underestimate the true impact of PAS administration by the 5 points the score was inflated.

There are a number of other results worth mentioning. While it is true that divided government at the time a program was created does not alter the management purpose grade, it does have a negative relationship with strategic planning, program management, and program results, as well as the total score. It has a significant relationship with strategic planning. While not entirely persuasive, these results are at least suggestive that the politics at the time of a programs creation might have lasting effects on the quality of program management. It is possible that long term effects on management have less to do with program clarity than on the contentiousness of the

---

[23] Lucier, Chuck, Rob Schuyt, and Eric Spiegel. 2003. "Deliver or Depart: CEO Succession 2002." *strategy+business* 31 (Summer 2003):32-45.

program over time. Program managers that must satisfy multiple bosses with diverging visions for the program probably serve for shorter periods, have a harder time doing long term planning, and suffer from increased micromanagement. Chang et al. (2000) find that political appointees have shorter tenures during periods of divided government.

There is some evidence that programs created under Democratic administrations were graded lower than programs created in Republican administrations. The coefficients in the last 4 models on Democratic president are all negative and two are significant at the 0.05 level. The coefficients on the Democratic Congress variable, however, are not significant and the coefficients are close to 0. Given this interesting finding, further research is necessary to determine the degree of politicization in the administration's grades. It illustrates one of the fears about performance measures that they are used as a cover for preexisting agendas or decisions.

Finally, agencies whose managers serve for fixed terms tend to have substantially higher management grades. The long tenure allows for long-term planning, lower turnover, and consistency at the top. When politicians are concerned about turnover, they have occasionally given managers fixed terms. The did so with the administrator of the National Nuclear Security Administration after repeated security lapses at the nation's weapons labs.

V. Discussion and Conclusion

We have shown that programs administered by political appointees get systematically lower management grades than programs administered by senior executives even when controlling for a variety of factors. A number of questions remain.

First, is it possible that the grades are corrupted enough to call into question the robustness of this finding? This is unlikely since the bias is unlikely to be correlated with whether a program is administered by a political appointee or not and because we have attempted to estimate models with proper controls. In addition, if the administration wanted to produce grades that were in its favor with regard to the political appointee/SES split, they would likely give programs run by political appointees higher grades than those run by senior executives. We found the opposite. This implies that if bias were a problem it would lead us to underestimate the true effect of political appointee management on program performance.

A second outstanding issue is what the implications of these finds are for public administration. The results certainly bolster the claims of the Volcker Commission and others who claim that programs administered by political appointees often suffer under poor management. A more robust discussion should begin on whether to give direct program management responsibility to senior executives rather than political appointees. The proper role for appointees may be at one level above the program management level. Programs undoubtedly would benefit from more continuity, clearer more stable goals, and better management. On the other hand, decreasing political appointee presence has drawbacks. In particular, it hurts the president's ability to control the administrative state. Modern presidents, probably unfairly, are held accountable for the functioning of the entire government and removing political appointees could make an administrative state that is already incredibly and increasingly difficult to manage even more difficult to control. One of the persistent pernicious features of the American administrative state is its fragmentation and the president at the top is the primary political actor with a national

constituency who voters can look to for integration. If we reduce the number of political appointees we will also have to roll back public expectations of presidential management.

This research also has implications for public administration in that it highlights productive avenues for future research. First, this research begs the question of why careerists run some programs and why appointees run others. Is there a political logic behind this determination? Second, there is substantial variation in the number and depth of political appointees in the various departments and agencies that administer federal programs. We know very little about why this variation exists or why patterns of "politicization" vary over time.

Finally, this research shows the value of the administration's management grades for public administration and political science research. In September 2003 the administration intends to release PART scores for 234 more programs. This new release will allow scholars to determine whether and to what extent performance measurement and budgets can be used to improve management performance. Use of performance measures and budgets to improve public management is a difficult task. In many cases government agencies are monopoly providers of goods. That is, if the federal government does not provide a good or service, there is no one else equipped to provide it. If a federal program gets a bad grade and a budget cut, this may just exacerbate the problem. If the Federal Election Commission, for example, is managed poorly and is not demonstrating results, cutting its budget is not likely to improve its performance. Some federal programs may perform poorly precisely because they are underfunded or understaffed. Budget and performance integration is a blunt tool to remedy what is a difficult and complex problem but it may work better for some programs than others.

This new data has the potential to allow us to theorize about when budget and performance integration can work the best and when performance measures should more appropriately point us toward other solutions to management deficiencies.

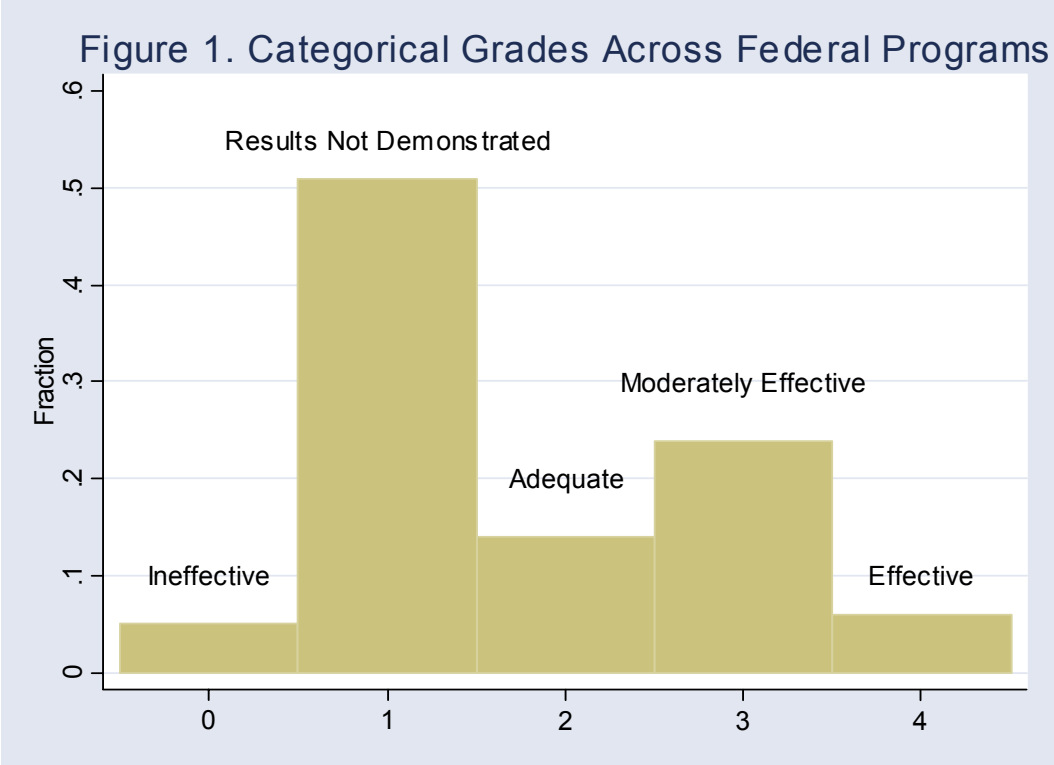
Figure 1. Categorical Grades Across Federal Programs
Fraction
0
.1
.2
.3
.4
.5
.6
Results Not Demonstrated
Moderately Effective
Adequate
Ineffective
Effective
0
1
2
3
4

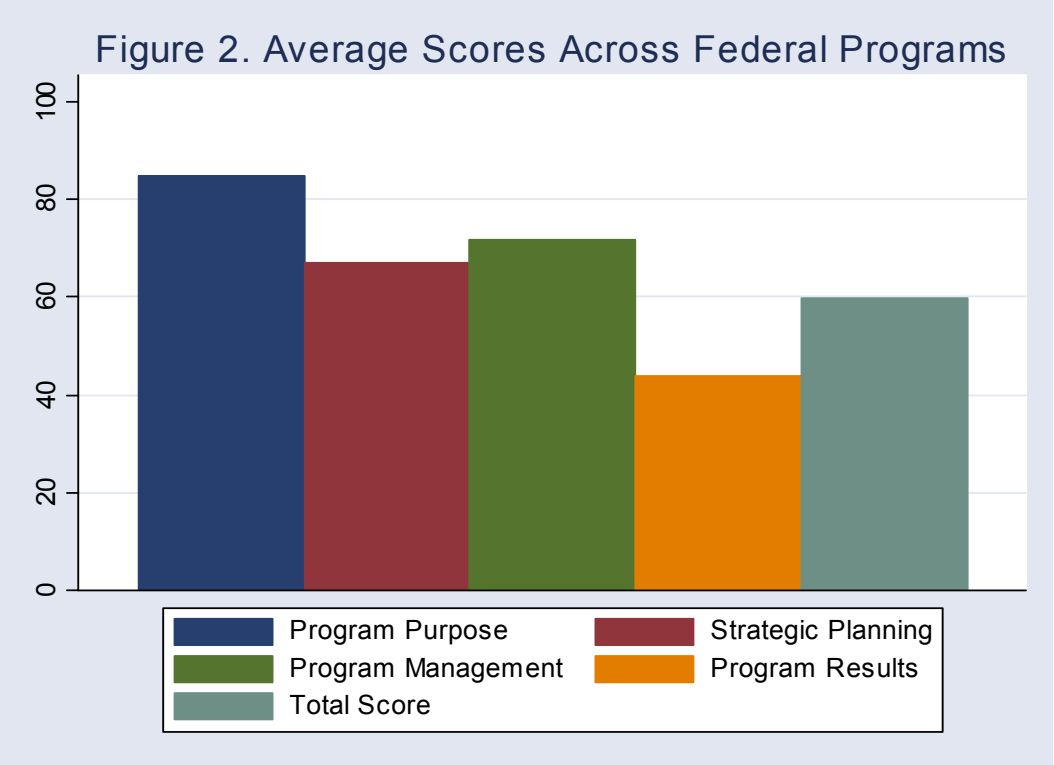
Figure 2. Average Scores Across Federal Programs
0
20
40
60
80
100
Program Purpose
Strategic Planning
Program Management
Program Results
Total Score

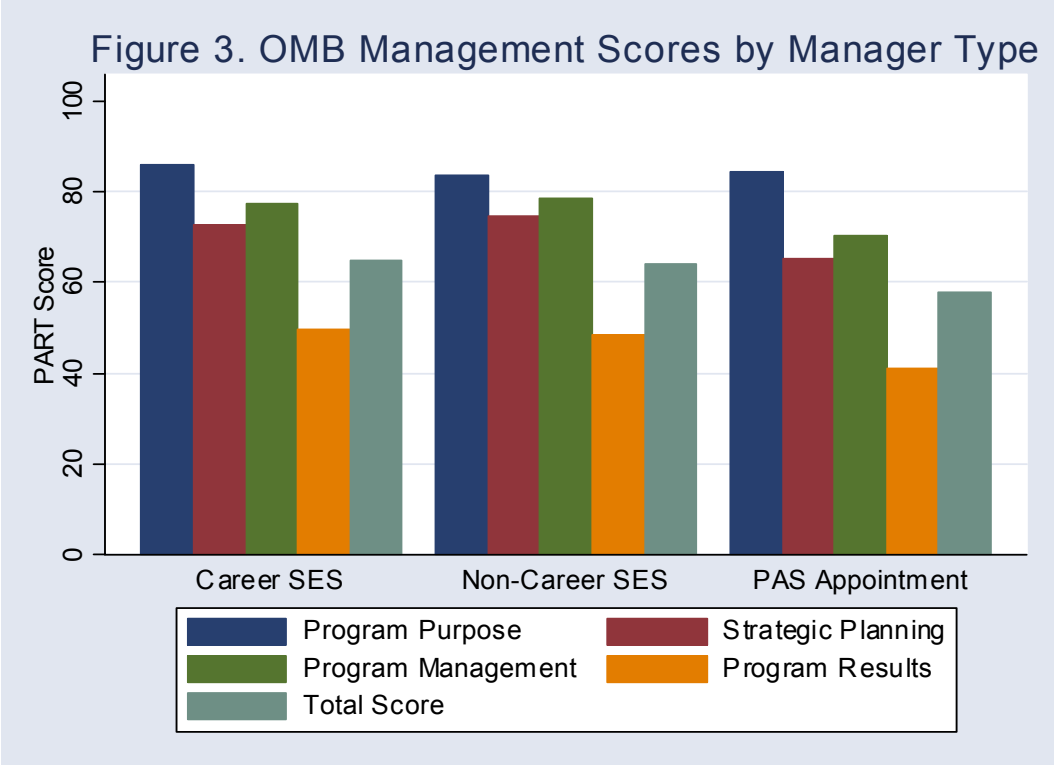
Figure 3. OMB Management Scores by Manager Type
PART Score
0
20
40
60
80
100
Career SES
Non-Career SES
PAS Appointment
Program Purpose
Strategic Planning
Program Management
Program Results
Total Score

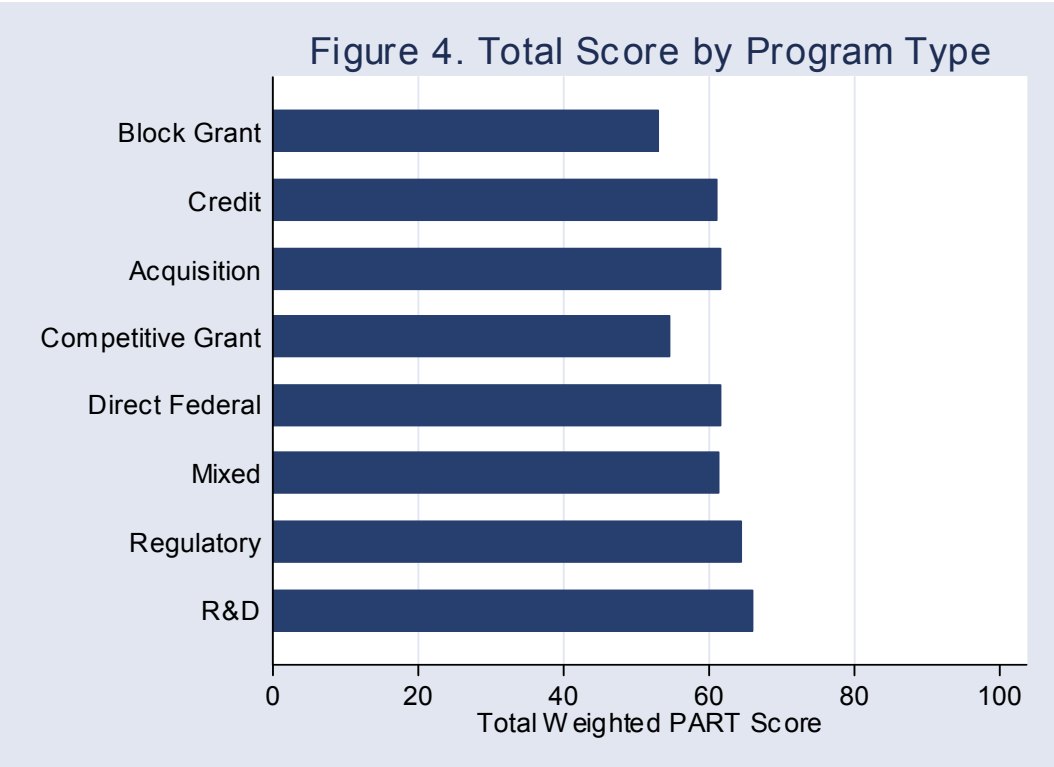
Figure 4. Total Score by Program Type
Block Grant
Credit
Acquisition
Competitive Grant
Direct Federal
Mixed
Regulatory
R&D
0
20
40
60
80
100
Total Weighted PART Score

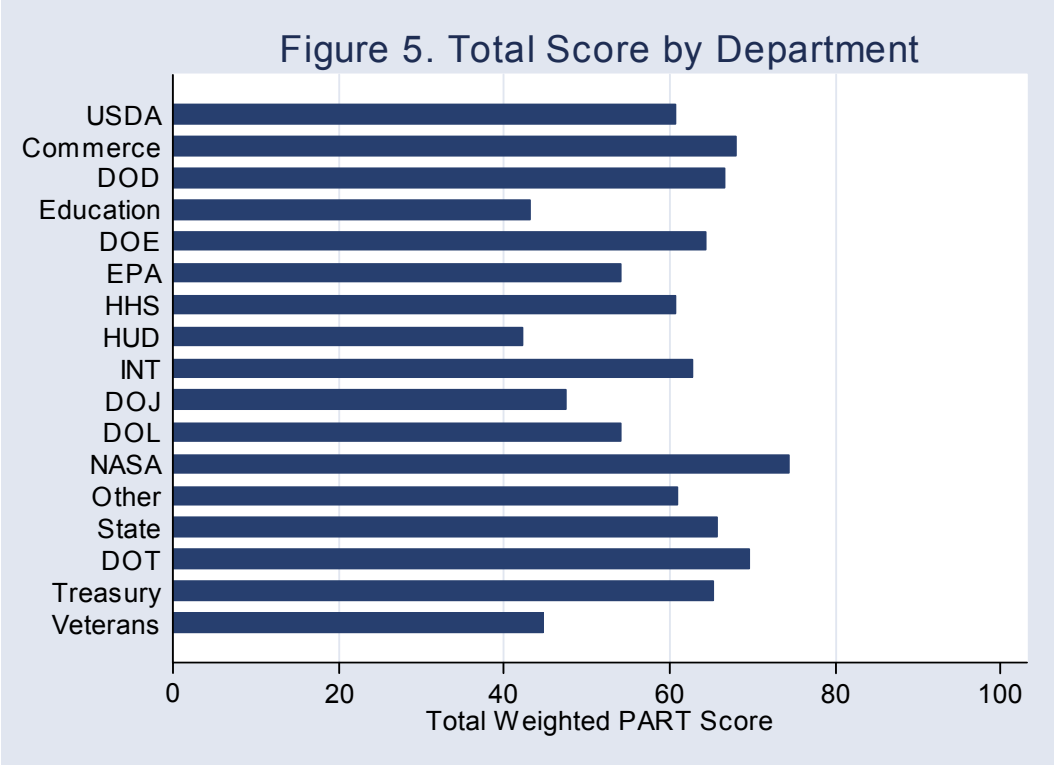
Figure 5. Total Score by Department
USDA
Commerce
DOD
Education
DOE
EPA
HHS
HUD
INT
DOJ
DOL
NASA
Other
State
DOT
Treasury
Veterans
0
20
40
60
80
100
Total Weighted PART Score

Table 1. Estimates of PART Score Grades for Federal Programs, FY 2004

	Program Purpose	Strategic Planning	Program Management	Program Results	Total Weighted	Categorical Grade
PAS Appointee (0,1)	4.00	-15.19**	-9.55**	-14.70**	-9.69**	-0.91**
	(6.72)	(4.27)	(3.73)	(4.74)	(2.85)	(0.25)
Divided Government (0,1)	2.40	-14.76**	-5.21	-5.45	-4.87	0.03
	(7.41)	(4.87)	(4.21)	(5.45)	(3.48)	(0.30)
Democratic President (0,1)	2.17	-14.06**	-5.30	-6.16	-5.54**	-0.38
	(5.98)	(3.86)	(3.36)	(4.37)	(2.70)	(0.24)
Democratic Congress (0,1)	3.48	2.23	-0.08	0.86	0.23	-0.06
	(5.13)	(3.39)	(2.95)	(3.81)	(2.49)	(0.20)
Fixed Term for Manager (0,1)	7.82	23.74**	6.45	29.98**	16.89**	1.50**
	(10.87)	(6.71)	(5.70)	(7.29)	(4.40)	(0.43)
Age of Program	0.09	-0.10	-0.05	0.08	0.02	0.007
	(0.13)	(0.08)	(0.07)	(0.09)	(0.06)	(0.006)
Constant	97.17**	74.72**	67.63**	49.04**	54.20**	--
	(25.62)	(15.99)	(13.69)	(17.89)	(20.10)	
N	183	186	186	186	186	186
Left-censored	0	3	0	14	0	0
N Right-censored	86	11	18	0	0	0
X^2, F (29 df)	57.90*	89.59**	78.22**	75.22**	3.72**	90.09**

Note: Fixed effects for program type and department omitted. ** significant at .05 level in two-tailed test of significance; *significant at the .10 level in two-tailed test of significance. Standard error regression in first four models is (26.24, 18.73, 16.37, 21.13). Since there are no censored observations for the total score, we estimate a linear regression with robust standard errors. Final column estimates are ordered probit estimates of overall grade (ineffective, results not demonstrated, adequate, moderately effective, effective). Cutpoint estimates are –1.86, 0.33, 0.79, 2.26.

Appendix A. Estimates of PART Score Grades for Federal Programs Including Budget Data, FY 2004

	Program Purpose	Strategic Planning	Program Management	Program Results	Total Weighted	Categorical Grade
PAS Appointee (0,1)	-0.43	-11.45**	-11.01**	-10.56*	-8.36**	-0.74**
	(7.54)	(4.53)	(4.17)	(5.43)	(3.59)	(0.31)
Divided Government (0,1)	11.21	-17.58**	-9.38**	-8.43	-5.65	0.07
	(8.55)	(5.17)	(4.68)	(6.30)	(4.16)	(0.35)
Democratic President (0,1)	7.19	-16.05**	-9.02**	-6.52	-5.51*	-0.51*
	(6.84)	(4.10)	(3.75)	(5.02)	(3.07)	(0.28)
Democratic Congress (0,1)	3.69	-1.61	0.81	-0.65	-0.63	-0.18
	(5.98)	(3.52)	(3.24)	(4.31)	(2.93)	(0.23)
Fixed Term for Manager (0,1)	-1.96	25.91**	16.07**	28.64**	20.40**	1.72**
	(11.98)	(7.73)	(7.02)	(9.13)	(5.05)	(0.52)
Age of Program	0.25*	-0.11	-0.11	0.03	0.01	0.007
	(0.13)	(0.09)	(0.08)	(0.10)	(0.07)	(0.007)
ln(Budget FY 2003)	0.90	1.02	-1.74	-0.00	-0.14	-0.08
	(2.00)	(1.17)	(1.09)	(1.44)	(0.94)	(0.07)
Constant	48.32	131.59**	95.57**	98.85**	74.56**	--
	(34.38)	(15.99)	(20.19)	(26.64)	(11.80)	
N	149	149	149	149	149	149
Left-censored	0	2	0	12	0	0
N Right-censored	74	9	12	0	0	0
X^2, F (29 df)	39.92**	91.85**	70.65**	63.84**	2.68**	62.86**

Note: Fixed effects for program type and department omitted. ** significant at .05 level in two-tailed test of significance; *significant at the .10 level in two-tailed test of significance. First model excludes program type dummies to facilitate estimation. Standard error regression in first four models is (27.42, 17.11, 15.90, 20.88). Since there are no censored observations for the total score, we estimate a linear regression with robust standard errors. Final column estimates are ordered probit estimates of overall grade (ineffective, results not demonstrated, adequate, moderately effective, effective). Cutpoint estimates are –4.68, -2.42, -1.87, -0.34.

References

Ban, Carolyn, and Patricia W. Ingraham. 1990. Short-Timers: Political Appointee Mobility and Its Impact on Political-Career Relations in the Reagan Administration. *Administration and Society* 22 (1):106-124.

Brauer, Carl. 1987. Tenure, Turnover, and Postgovernment Employment Trends of Presidential Appointees. In *The In-and-Outers: Presidential Appointees and Transient Government in Washington*, edited by G. C. Mackenzie. Baltimore: The Johns Hopkins University Press.

Chang, Kelly, David E. Lewis, and Nolan M. McCarty. 2000. Turnover Among Political Appointees. Paper presented at annual meeting of the American Political Science Association, Washington, DC.

Chang, Kelly, Stuart V. Jordan, David E. Lewis, Nolan M. McCarty. 2003. The Tenure of Political Appointees. Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL.

Congress and the Nation, v. 8, 1989-1992. Washington, DC: CQ Press.

Daalder, Ivo H., I.M. Destler, James M. Lindsay, Paul C. Light, Robert E. Litan, Michael E. O'Hanlon, Peter R. Orszag, and James B. Steinberg. 2002. Assessing the Department of Homeland Security. Washington, DC: Brookings Institution.

General Accounting Office. 1994. *Political Appointees: Turnover Rates in Executive Schedule Positions Requiring Senate Confirmation*. Washington, DC: GAO.

Heclo, Hugh. 1977. *A government of strangers: executive politics in Washington*. Washington: Brookings Institution.

Joyce, Philip G. 1990. An Analysis of the Factors Affecting the Employment Tenure of Federal Political Executives. *Administration and Society* 22 (1):127-45.

Lewis, David E. 2003. *Presidents and the politics of agency design: political insulation in the United States Government Bureaucracy, 1946-1997*. Stanford, CA: Stanford University Press.

Light, Paul. 1992. *Forging Legislation*. New York, NY: W.W. Norton.

Light, Paul. 1995. *Thickening Government: Federal Hierarchy and the Diffusion of Accountability*. Washington, DC: Brookings.

Lucier, Chuck, Rob Schuyt, and Eric Spiegel. 2003. "Deliver or Depart: CEO Succession 2002." *strategy+business* 31 (Summer 2003):32-45.

Mackenzie, G. Calvin, ed. 1987. *The In-and-Outers: Presidential Appointees and Transient Government in Washington*. Baltimore: The Johns Hopkins University Press.

Moe, Terry M. 1982. Regulatory Performance and Presidential Administration. *American Journal of Political Science* 26 (2):197-224.

Moe, Terry M. 1985. The Politicized Presidency. In *The New Direction in American Politics*, edited by J. E. Chubb and P. E. Peterson. Washington, DC: Brookings.

Moe, Terry M. 1989. The Politics of Bureaucratic Structure. In *Can the Government Govern?*, edited by J. E. Chubb and P. E. Peterson. Washington, DC: Brookings.

Nathan, Richard P. 1975. *The Plot that Failed: Nixon and the Administrative Presidency*. New York: John Wiley.

Nathan, Richard P. 1983. *The Administrative Presidency*. New York: John Wiley.

National Commission on the Public Service. 2003. *Urgent Business for America: Revitalizing the Federal Government for the 21st Century*. Washington, DC: Brookings.

Stanley, David T., Dean E. Mann, and Jameson W. Doig. 1967. *Men Who Govern: A Biographical Profile of Federal Political Executives*. Washington, DC: Brookings.

U.S. Congress. Senate. Committee on Governmental Affairs. *Policy and Supporting Positions.* 106th Congress, 2nd Sess.

U.S. Office of Management and Budget. *Instructions for the Program Assessment Ratings Tool*. Washington, DC, July 12, 2002.

U.S. Office of Management and Budget. *M-02-10. Memorandum for Heads of Executive Departments and Agencies*. Washington, DC, July 16, 2002.

U.S. Office of Management and Budget. 2003. *Budget of the United States Government FY 2004: Performance Management and Assessments*. Washington, DC: U.S. Government Printing Office.

U.S. President's Foreign Intelligence Advisory Board. 1999. *Science at its Best, Security at its Worst: A Report on Security Problems at the U.S. Department of Energy*. June, 1999.