Political Appointees and the Competence of Federal Program Management

John B. Gilmour
College of William & Mary

David E. Lewis
Princeton University

In this article, we use the Bush administration's management grades to analyze whether programs administered by senior executives are better managed than those administered by political appointees requiring Senate confirmation. 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 review of the proper role of political appointees in federal program management should be considered.

Keywords: *public administration; political appointees; Senior Executive Service (SES); government management; Program Assessment Rating Tool (PART); executive branch*

T he implementation of public policies is a complex process with numerous potential failures along the way from articulation of goals to the delivery of a product. Slippage between what politicians want and what administrative agencies actually deliver can be caused by a variety of factors, including resource constraints, political disagreements, difficulty observing outcomes, task size, and the complexity of joint action (e.g., Bardach, 1977; Pressman & Wildavsky, 1974; Tendler, 1997; Wilson, 1989). Management

capacity of program administrators, or lack thereof, can also account for the failure of actual outputs of programs to match initial promise.

Program managers can interfere with effective implementation for multiple reasons. First, they may not share the goals of their political superiors and fail to carry out policies as intended. Second, they may be ineffective managers. Fear of bureaucratic sabotage has led presidents, at least since Nixon, to pursue an administrative presidency whereby they seek to control policy implementation through an appointment strategy (Moe, 1982, 1985; Nathan, 1975, 1983). This has led both to an increasing number of managers and an increasing number of political appointees among managers in the executive branch (Light, 1995).

Ironically, the approach to improving management by imposing political control also poses a threat to effective management because the individuals appointed to political positions in the bureaucracy may not be good managers. Public managers who understand exactly the policy outcomes politicians want will vary in their individual ability to translate these goals into reality. If appointees are less capable as managers than careerists, the appointments strategy may undermine effective management.

This leads to an important question: Are political appointees as competent as bureaucratic careerists? This question has received much discussion among public administrators but little among political scientists. Yet because it is ultimately a question of the relationship between institutional arrangements and policy outputs, it is an important issue for political science.

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 (National Commission on the Public Service, 2003). 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 (see Brauer, 1987; Chang, Jordan, Lewis, & McCarty, 2003; Chang, Lewis, & McCarty, 2000; General Accounting Office, 1994; Mackenzie, 1987; Stanley, Mann, & Doig, 1967). 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 interagency 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. 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 to 600 presidential appointees requiring Senate confirmation (PAS positions).[1] According to the Office of Management and Budget, political appointees administer some federal programs and senior executives, either career or noncareer, run the vast majority of the rest.[2]

Despite claims that programs run by political appointees are worse off than programs run by managers drawn from the Senior Executive Service (SES), there is no consensus that this is the case. Donahue (2003) examines first-year appointments following a partisan shift and finds that appointees are better educated now than before, and they are more likely now to have served in government prior to their appointment. This suggests that the quality of program management by appointees may be improving. Maranto (1998) argues that political appointees stay longer than widely believed, are generally perceived by career federal executives as competent, and that presidents should reject calls to reduce the number of political appointees. Michaels (1997) reports that 53% of George H. W. Bush appointees had advanced degrees, 49% had previously managed budgets larger than $1 million, and 65% were older than the age of 50. Each suggests that political appointees might be more competent than widely perceived.

Others are more dubious of the relative quality of management by careerists. Bilmes and Neal (2003) argue that the antiquated federal personnel system hurts recruitment, retention, and ultimately the quality of the career service. They suggest that the problem is only going to get worse as the aging workforce prepares to retire. Bok (2003) notes that career civil service systems can promote secrecy, inertia, elitism, and insularity that hamper performance and argues that the system of having several thousand appointees "counteracts inertia, ensures an influx of new ideas, and keeps government in touch with a variety of interested groups and constituencies" (p. 265).

Unfortunately, there is little empirical evidence to arbitrate between competing views of political appointee management. One difficulty with assessing comparative management competence is disagreement about what competent management is. For administration officials, for example, an important component of good management is responsiveness to the policy program of the chief executive. This aspect of good management is less important to members of Congress or other public administration scholars.

Comparing the management quality of appointee-run versus careerist-run programs is also complicated by the fact that it is extremely difficulty to measure comparative management quality of any type. 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 because 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 provide a unique opportunity for scholars to study federal management. The Office of Management and Budget designed and publicly vetted a series of measures to assess the quality of federal program management. They settled on a definition of the components of good management and adjusted the measures of federal program management to be applicable across programs, taking into account unique differences in program types. The stated intention of the administration is to use these management grades to determine increases or decreases in each program's budget. They reflect the most ambitious attempt in history to measure program management quality.

In this article, we use these grades to compare the management of federal programs by political appointees and managers from the SES. We deal with concerns about the validity of these measures and their usefulness for comparing appointee versus senior executive management in a meaningful way in the first section of the article. In the second section, we explain why 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 article, 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 noncareer 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.

Can Program Assessment Rating Tool (Part) Scores Tell Us Something About Management Quality?

One concern about the management scores is whether they can be used reliably to make inferences about the comparative management quality of federal programs. This is a serious concern, and we address it in two ways. First, we describe the grades and how they are constructed to show that they probably capture some of the variance in the quality of federal program management. Second, we explain how even if the grades are flawed, they can tell us something useful about comparative management quality. Specifically, we explain why the possible politicization or inaccuracy in the grades does not prevent us from making conclusions about the comparative management quality of programs run by appointees or senior executives.

The Bush administration's management grading scheme is relatively straightforward. It was designed by the Office of Management and Budget (OMB) in consultation with the President's Management Council, the Performance Measure Advisory Council, a group from the President's Council on Integrity and Efficiency, the National Academy of Public Administration, and interested members of Congress (U.S. Office of Management and Budget, 2003b). It has been subjected to public comment from agency officials, public management specialists, and interested observers since its introduction in 2002. It includes numerical grades from 0 to 100 in four categories and a final, total, weighted numerical management grade. The four categories with their purposes are (U.S. Office of Management and Budget, 2002):[3]

1. Program purpose and 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 on answers to a series of yes or 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?" Other questions used to evaluate management quality include

- Are federal managers and program partners (grantees, subgrantees, contractors, etc.) held accountable for cost, schedule, and performance results?
- Has the program taken meaningful steps to address its management deficiencies?
- 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?
- Does the program (including program partners) achieve its annual performance goals?

For these 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.

Demonstrating the validity of the PART scores is a challenge because there is no available systematic alternative measure of performance and management with which we can compare PART scores. If there were, we could do a test of convergent validity, which is, in essence, a test of whether a measure correlates appropriately with another measure that it approximates. According to the literature on validity, in the absence of a standard for comparison, the most appropriate tests of validity are face validity and content validity (Carmines & Zeller, 1979). Face validity means that when examining the measure, it makes sense on its face. Content validity means that the measure addresses all aspects of the concept it purports to assess. Assessing the validity of PART scores requires, then, examining the questions and observing whether they appear to relate reasonably to the quality of program management. They are reproduced here as an appendix (see Appendix A), and we encourage readers to examine them closely. Lacking an alternative standard for comparison, the best we can do in establishing the validity of the PART scores is to consider the process by which they were produced and whether the questions that comprise the questionnaire reasonably relate to management effectiveness and performance. This is necessarily a subjective judgment, but the questions do seem to relate to appropriate aspects of performance and management, and, thus, we believe that the PART scores pass, at a minimum, the tests of both face and content validity.

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 so-called Democratic programs such as 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. These are serious concerns, but, even if true, they do not eliminate the usefulness of the grades. Provided that the grades correlate somewhat with true management quality and that any errors are not correlated with the appointment status of program managers, we can use the grades to compare the management quality of programs run by appointees and senior executives.

It seems unlikely that grades and true management quality would be completely uncorrelated. If one program were objectively horribly managed and another was managed extremely well, would this be picked up in grades of the two programs? Probably so. 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 PART scorers detect this? The answer to those simple questions is probably yes, even if there is error, bias, or favoritism in assigning grades. Even a very imperfect system of measuring performance would probably be able to discern gross differences in program management, although small differences would be harder to discern.

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 nonrandom and not only nonrandom 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 isolate 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? Although 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), and so forth.

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 inclusion programs that were recommended for review by interested parties and programs coming up for reauthorization.[4] The administration is adding another 20% of federal programs into the grading scheme for the fiscal year (FY) 2005 budget and suggesting that their intent is to grade all federal programs during 5 years.

In sum, for management grades to be useful in comparing career and appointed managers, the grades must correlate somewhat with true management quality. The correlation need not be very high. A significant amount of error in the grades is tolerable, provided that the error does not correlate with the variable of greatest interest here: whether a manager is career or appointed. Although there may be bias in favor of Republican programs—or error from rating programs incorrectly—we doubt that these errors correlate with the key theoretical variable. Even if imperfect, OMB's management scores can be helpful in assessing management quality.

The Advantage of
Senior Executive-Run Programs

There are several points in the politician to program manager to program management chain where problems can arise. These problems include different preferences or priorities, monitoring difficulties, or a lack of competence at one or more levels in the chain. Each can lead to slippage between what the politicians desire and what is actually produced. Senior executives have advantages over political appointees in avoiding each of these problems.

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 because of 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. 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 (e.g., Dodd & Schott, 1979; McCubbins, Noll, & Weingast, 1987; McCubbins & Schwartz, 1984). 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 underequipped. 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 and 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. Although 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 panoply of stakeholders whose support is necessary for crafting 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 during 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 (e.g., Ban & 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 (President's Foreign Intelligence Advisory Board, 1999).

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.

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 SES are more likely to have these skills, and programs run by senior executives are more likely to have managers of this type in succession.

It should also be noted, however, that perceptions of management competence are invariably connected to policy preferences. Presidents have sought more political appointees precisely because appointees are more responsive to presidential direction. Presidents do not want program managers setting policy in line with the preferences of other stakeholders like clients, beneficiaries, interest groups, or Congress except to the extent that it helps the administration get what it wants. Similarly, Congress wants program managers to implement policy in line with program purpose as embedded in law and communicated by congressional committees. They do not want program managers being sensitive to the interests of presidents or other stakeholders except to the extent that this is consistent with their preferences. This is one reason why the Bush administration's management grades present a unique opportunity to evaluate comparative management quality. Because the management scores are designed by the administration, if bias exists in the definition of management quality, it is probably bias in favor of political appointees. The administration's definition of management competence, if anything, is overly sensitive to the extent to which program managers reflect the interests of the administration. Appointees are likely to have a systematic advantage over senior executives in such a grading scheme. As such, any evidence that senior executive–run programs get higher management grades is all the more convincing because the grades are probably biased to favor those managers that share the administration's preferences (i.e., political appointees).

Figure 1
**Categorical Grades Across Federal Programs**
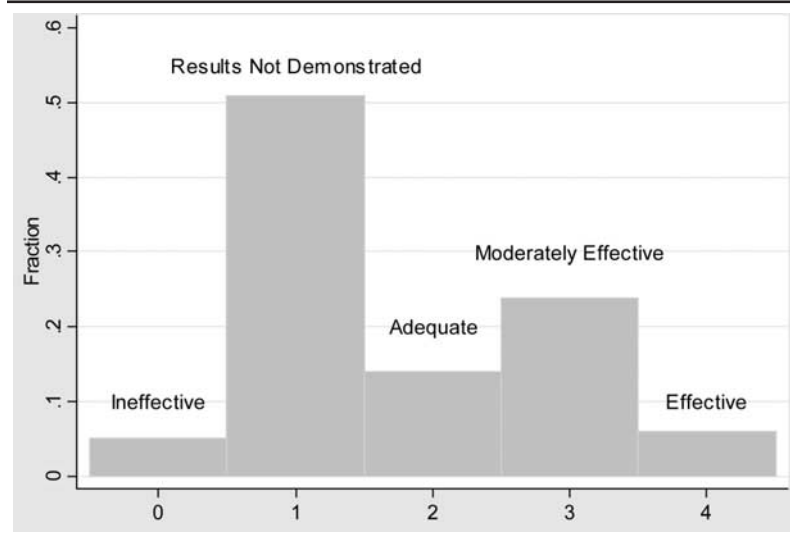
Data, Variables, and Methods

Our expectation is that senior executive–run programs 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 (0 to 100) 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. More than half were graded either "ineffective" or "results not demonstrated." The categorical grade and the total numerical grade were determined by raw scores in four areas that were weighted and summed. We graph the average scores in all four management areas along with the total management score in Figure 2. The lowest scoring program overall received an 11 and the maximum was a 93 out of 100. The average score was 59.76 ($SD = 16.78$). The disability compensation program in the Department of Veterans Affairs (11/100) and

Figure 2
**Average Scores Across Federal Programs**
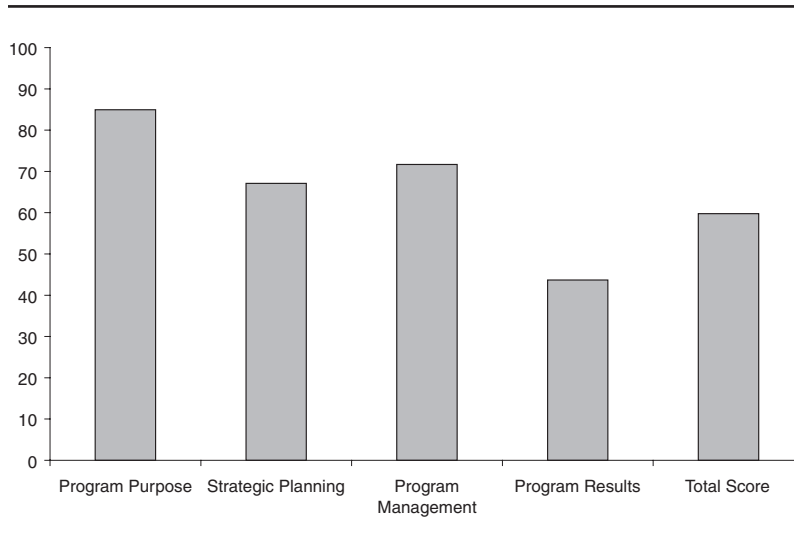
the Individuals With Disabilities Education Act Preschool Grants Program in the Department of Education (19/100) were the two lowest scoring programs. Broken down by category, the average raw score for program purpose and design was 84.9 ($SD = 18.36$) with a minimum of 20 and a maximum of 100. The mean 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 ($SD = 24.3$).

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 surprising, program results had the lowest average score (43.7) because this score is partly determined by success in the other three categories.

PAS or SES

To determine whether a program was administered by a political appointee or senior executive, we consulted OMB worksheets that list the bureau administering each program (U.S. Office of Management and Budget, 2003a). We then looked up the bureau in what is commonly referred to as

Figure 3
Office of Management and Budget
Management Scores by Manager Type
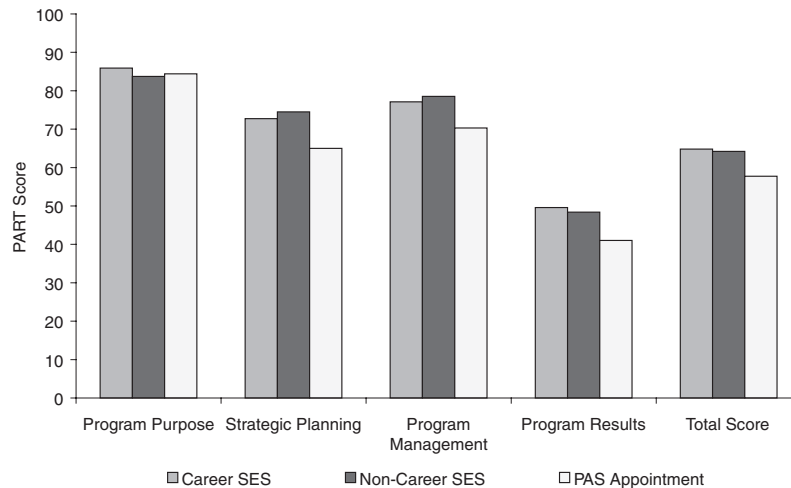
the 2000 *Plum Book* to determine the appointment status of the bureau chief (U.S. Senate, 2000). One complication with this approach is that the career/ noncareer status of SES bureau chiefs can change.[5] Many programs will have been run by both career and noncareer SES managers at different points in their history. This makes it difficult to distinguish career SES–run programs from noncareer SES–run programs and consequently difficult to asses the relative management quality of career and noncareer managers. We report data on noncareer SES versus career SES and political appointees with this caveat.

Of 214 programs for which data were available, 165 (77%) were administered by bureaus headed by political appointees. Of the remaining 49 programs, career employees of the SES administered 20, and noncareer 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 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 three cat-

egories 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 to 8 points and statistically significant in difference of mean tests ($p < .05$).

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 its own political appointees higher than career employees. The results, however, suggest that even with this potential bias, SES-run agencies get higher grades.[6]

It is worth noting that although there is a significant difference between political appointments that require Senate confirmation and SES managers, there is virtually no difference between career and noncareer SES managers in the scores. This may reflect the fact that noncareer 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 versus noncareer SES reflects the fact that career status of SES program managers changes depending on the choice of the current administration.

Regression Analysis

There are a number of possible objections to the relationship described above between SES and 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 (see Figure 4).[7]

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 and 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 (see Figure 5).

Figure 4
Total Score by Program Type
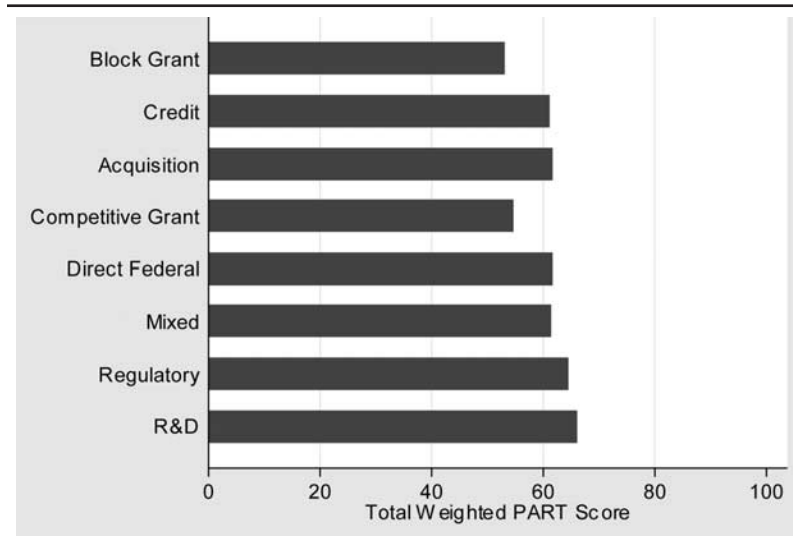
To address these possibilities, we turn to regression analysis, which will allow us to control for some of the possible confounding factors. To account for the differences in program type, we include dummy variables for each type—competitive grant, block/formula grant, regulatory, capital assets and service acquisition, credit, direct federal, and research and development programs (U.S. Office of Management and Budget, 2002). We also include indicators (0,1) for each agency that houses a program.

Some of the difficulty or ease of managing a program may be because of 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.

Figure 5
Total Score by Department
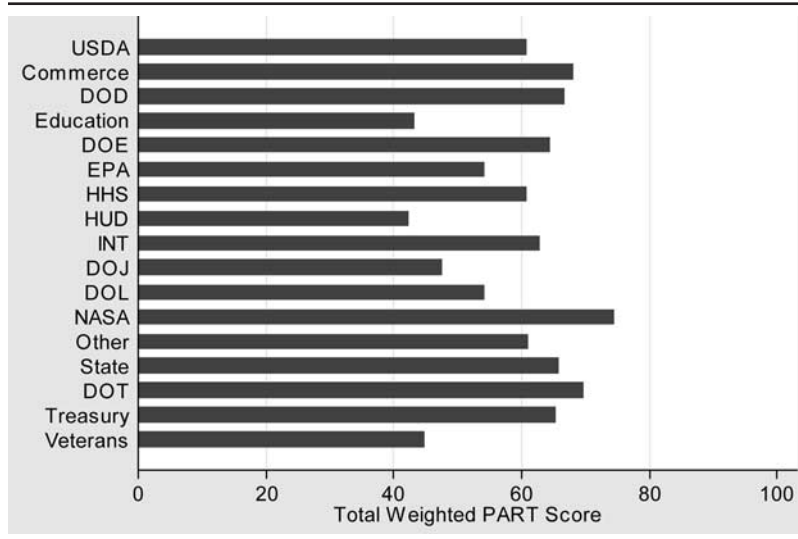
To account for the politicization that may be present in grading, we include indicators for the presence of a Democratic president (0,1) and a Democratic Congress (House and Senate, 0,1). If the grading is really politicized, programs created under Democratic presidents and Democratic congresses should 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 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.[8] To account for this, we include an indicator for whether the bureau chief in question has a fixed term (U.S. Senate, 2000). 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 (U.S. Office of Management and Budget, 2003a). 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 B. Finally,

some programs are much older than other programs.[9] The patent program and the trademark program, for example, trace their histories back at least 200 years. Our expectation is that older programs should be managed better because of experience, institutional memory, and learning throughout 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 (i.e., 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.[10] 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 model of outcomes.[11] We note, however, that the type of model makes little substantive difference for the estimates.

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 political appointees are graded significantly worse on different dimensions of public management. Table 1 includes six columns, one for each management component, a column for the total numerical score, and a final column modeling the categorical grade.

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

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

	Dependent Variables					
Independent Variables	Program Purpose	Strategic Planning	Program Management	Program Results	Total Weighted	Categorical Grade
PAS appointee (0,1)	4.00 (6.72)	−15.19** (4.27)	−9.55** (3.73)	−14.70** (4.74)	−9.69** (2.85)	−0.91** (0.25)
Divided government (0,1)	2.40 (7.41)	−14.76** (4.87)	−5.21 (4.21)	−5.45 (5.45)	−4.87 (3.48)	0.03 (0.30)
Democratic president (0,1)	2.17 (5.98)	−14.06** (3.86)	−5.30 (3.36)	−6.16 (4.37)	−5.54** (2.70)	−0.38 (0.24)
Democratic Congress (0,1)	3.48 (5.13)	2.23 (3.39)	−0.08 (2.95)	0.86 (3.81)	0.23 (2.49)	−0.06 (0.20)
Fixed term for manager (0,1)	7.82 (10.87)	23.74** (6.71)	6.45 (5.70)	29.98** (7.29)	16.89** (4.40)	1.50** (0.43)
Age of program	0.09 (0.13)	−0.10 (0.08)	−0.05 (0.07)	0.08 (0.09)	0.02 (0.06)	0.007 (0.006)
Constant	97.17** (25.62)	74.72** (15.99)	67.63** (13.69)	49.04** (17.89)	54.20** (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: PART = Program Assessment Rating Tool; FY = fiscal year; PAS = presidential appointee requiring Senate confirmation. Fixed effects for program type and department were omitted. Standard error regression in the first four models is 26.24, 18.73, 16.37, 21.13. Because 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.

$*p < .10.$ $**p < .05.$

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 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. Because of 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 throughout time. We have done some additional analyses that suggest this might be the case, but the results are not very robust.[12]

After taking into account program purpose and design, the estimates confirm much of what we expected. In particular, the coefficient on PAS appointee is significant and negative in the five 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 to 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 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. Throughout 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 more than 2 to 3 years when the average corporate CEO serves from 5 to 7 years (Lucier, Schuyt, & Spiegel, 2003). 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. Because 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. Although 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. Although not entirely persuasive, these results are at least suggestive that the politics at the time of a program's 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 program throughout time. Program managers who 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) found 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 four 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. They did so with the administrator of the National Nuclear Security Administration after repeated security lapses at the nation's weapons labs.

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 because 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, it 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 findings are for political science and 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. Decreasing the number of appointees will also diminish the influx of new ideas and energy in government. With the improvement in the quality of appointees and the possible decline in the quality of the merit system, we should proceed cautiously instead of recklessly cutting numbers of appointees.[13]

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 throughout time.

Finally, this research shows the value of the administration's management grades for public administration and political science research. In January 2004, 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.

Appendix A
Key Questions Used in the PART Grading System by Management Area (U.S. Office of Management and Budget, 2003a)

Program Purpose and Design

- Is the program purpose clear?
- Does the program address a specific interest, problem, or need?
- Is the program designed to have a significant impact in addressing the interest, problem, or need?
- Is the program designed to make a unique contribution in addressing the interest, problem, or need (i.e., is not needlessly redundant of any other federal, state, local, or private effort)?
- Is the program optimally designed to address the national interest, problem, or need?

Strategic Planning

- 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?
- Does the program have a limited number of annual performance goals that demonstrate progress toward achieving the long-term goals?

- Do all partners (grantees, subgrantees, contractors, etc.) support program planning efforts by committing to the annual and/or long-term goals of the program?
- Does the program collaborate and coordinate effectively with related programs that share similar goals and objectives?
- Are independent and quality evaluations of sufficient scope conducted on a regular basis or as needed to fill gaps in performance information to support program improvements and evaluate effectiveness?
- Is the program budget aligned with the program goals in such a way that the impact of funding, policy, and legislative changes on performance is readily known?
- Has the program taken meaningful steps to address its strategic planning deficiencies?

Program Management

- Does the agency regularly collect timely and credible performance information, including information from key program partners, and use it to manage the program and improve performance?
- Are federal managers and program partners (grantees, subgrantees, contractors, etc.) held accountable for cost, schedule, and performance results?
- Are all funds (federal and partners) obligated in a timely manner and spent for the intended purpose?
- Does the program have incentives and procedures (e.g., competitive sourcing/cost comparisons, IT improvements) to measure and achieve efficiencies and cost-effectiveness in program execution?

Program Results

- Has the program demonstrated adequate progress in achieving its long-term outcome goal(s)?
- Does the program (including program partners) achieve its annual performance goals?
- Does the program demonstrate improved efficiencies and cost-effectiveness in achieving program goals each year?
- Does the performance of this program compare favorably to other programs with similar purpose and goals?
- Do independent and quality evaluations of this program indicate that the program is effective and achieving results?

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

Independent Variable	Dependent Variables					
	Program Purpose	Strategic Planning	Program Management	Program Results	Total Weighted	Categorical Grade
PAS appointee (0,1)	-0.43 (7.54)	-11.45** (4.53)	-11.01** (4.17)	-10.56* (5.43)	-8.36** (3.59)	-0.74** (0.31)
Divided government (0,1)	11.21 (8.55)	-17.58** (5.17)	-9.38** (4.68)	-8.43 (6.30)	-5.65 (4.16)	0.07 (0.35)
Democratic president (0,1)	7.19 (6.84)	-16.05** (4.10)	-9.02** (3.75)	-6.52 (5.02)	-5.51* (3.07)	-0.51* (0.28)
Democratic Congress (0,1)	3.69 (5.98)	-1.61 (3.52)	0.81 (3.24)	-0.65 (4.31)	-0.63 (2.93)	-0.18 (0.23)
Fixed term for manager (0,1)	-1.96 (11.98)	25.91** (7.73)	16.07** (7.02)	28.64** (9.13)	20.40** (5.05)	1.72** (0.52)
Age of program	0.25* (0.13)	-0.11 (0.09)	-0.11 (0.08)	0.03 (0.10)	0.01 (0.07)	0.007 (0.007)
ln (budget FY 2003)	0.90 (2.00)	1.02 (1.17)	-1.74 (1.09)	-0.00 (1.44)	-0.14 (0.94)	-0.08 (0.07)
Constant	48.32 (34.38)	131.59** (15.99)	95.57** (20.19)	98.85** (26.64)	74.56** (11.80)	—
N	149	149	149	149	149	149
Left-censored	0	2	0	0	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: PART = Program Assessment Rating Tool; FY = fiscal year; PAS = presidential appointee requiring Senate confirmation. Fixed effects for program type and department were omitted. The first model excludes program-type dummies to facilitate estimation. Standard error regression in the first four models is 27.42, 17.11, 15.90, 20.88. Because 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.

$*p < .10$. $**p < .05$.

Notes

1. 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 presidential appointees requiring Senate confirmation (PAS) positions but includes minor appointments to various small commissions, ambassadorships, and Justice Department positions that are Senate-confirmed by tradition such as U.S. marshals and U.S. attorneys.

2. According to the Office of Management and Budget (OMB), in the sample of federal programs that is the subject of this study, political appointees administer 165 (77%) of the programs. Career Senior Executive Service (SES) managers direct 20 (9%), noncareer managers run 28 (13%), and one program was administered by someone excepted from the competitive service by statute (Student Aid Administration).

3. See also U.S. Office of Management and Budget (2003a).

4. As described by Marcus Peacock (2003), program associate director for the OMB.

5. 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.

6. One potential source of bias is the appointee/senior executive status of those assigning the PART scores. If the examiners were senior executives, their loyalty to the SES might lead them to systematically grade the management of SES-run programs higher. The examiners, however, are civil service employees, sometimes relatively junior civil service employees of the OMB. In addition, even if the examiners were senior executives, it is unlikely that they would systematically bias the grades of SES-run programs upward. OMB employees have equally strong cross-pressures to please the administration, and there is very little professional consciousness in the SES or the civil service. Employees are dramatically more likely to identify with the agency they work for than the SES or the civil service system. This is a product of the diverse training backgrounds of employees in the United States versus other countries, the porous nature of the civil service system in the United States, weak and fragmented government employee unions, and the fact that career employees are dramatically more likely to work in one agency for a long time in the United States than civil servants are in other countries.

7. Our finding that programs run by appointees get lower grades may be threatened if programs that are inherently difficult to manage are assigned to appointees. If so, the fact that programs administered by political appointees get lower grades is an artifact of bias in the assignment process rather than something intrinsic to management by political appointees versus management by careerists. We have several responses to this possibility. First, it may be that the real bias works the other way and leads us to underestimate the true negative impact of appointee management. Congress and the president may actually be more likely to give responsibility for managing difficult programs to senior executives precisely because the programs are challenging to manage. Agencies such as NASA and the EPA have relatively low percentages of political appointees. Second, limited available evidence in our data does not support the idea that appointees are given responsibility for difficult programs. We thought, for example, that newer programs might be harder to manage because with new programs there is a process of learning how to run the program well and how to accomplish program goals. We might expect to see new programs being given to appointees to manage. However, in a comparison of programs that are less than 10 years old with older programs, new programs have a higher percentage of careerists (32%) than

older programs (21%). Finally, although one might hope that we could use econometric techniques to uncover evidence of bias in assignment of appointees to harder programs, this is not appropriate if there is no evidence of bias and difficult with unclear and potentially multiple processes determining the bias. As suggested above, existing social-science theory does not provide much guidance as to how to model the PAS/SES choice in a two-stage model, and finding appropriate instruments (regressors that are correlated with PAS/SES choice but not PART scores) is both necessary and extremely difficult.

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

9. 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 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 Congressional Quarterly *Federal Regulatory Directory* and other guides to federal programs.

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

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

12. 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.

13. For an interesting and balanced look at the status, strengths, and weaknesses of the current personnel system in the United States, see Donahue and Nye (2003) generally and Bok (2003) specifically. For a good discussion of the previous government experience and education of appointees throughout time, see Donahue (2003). For a critique of the federal personnel management system, see Bilmes and Neal (2003).

References

Ban, C., & Ingraham, P. W. (1990). Short-timers: Political appointee mobility and its impact on political-career relations in the Reagan administration. *Administration and Society*, *22*, 106-124.

Bardach, E. (1977). *The implementation game: What happens after a bill becomes a law?* Cambridge, MA: MIT Press.

Bilmes, L. J., & Neal, J. R. (2003). The people factor: Human resources reform in government. In J. D. Donahue & J. S. Nye Jr. (Eds.), *For the people: Can we fix public service?* (pp. 113-133). Washington, DC: Brookings Institution.

Bok, D. (2003). Government personnel policy in comparative perspective. In J. D. Donahue & J. S. Nye Jr. (Eds.), *For the people: Can we fix public service?* (pp. 255-272). Washington, DC: Brookings Institution.

Brauer, C. (1987). Tenure, turnover, and postgovernment employment trends of presidential appointees. In G. C. Mackenzie (Ed.), *The in-and-outers: Presidential appointees and transient government in Washington* (pp. 174-194). Baltimore: Johns Hopkins University Press.

Carmines, E. G., & Zeller, R. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

Chang, K., Jordan, S. V., Lewis, D. E., & McCarty, N. M. (2003, April). *The tenure of political appointees*. Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL.

Chang, K., Lewis, D. E., & McCarty, N. M. (2000, September). *Turnover among political appointees*. Paper presented at the annual meeting of the American Political Science Association, Washington, DC.

Dodd, L. C., & Schott, R. L. (1979). *Congress and the administrative state: Viewpoints on American politics*. New York: Wiley.

Donahue, J. D. (2003). In-and-outers: Up or down? In J. D. Donahue & J. S. Nye Jr. (Eds.), *For the people: Can we fix public service?* (pp. 55-71). Washington, DC: Brookings Institution.

Donahue, J. D., & Nye, J. S., Jr. (Eds.). (2003). *For the people: Can we fix public service?* Washington, DC: Brookings Institution.

General Accounting Office. (1994). *Political appointees: Turnover rates in executive schedule positions requiring senate confirmation*. Washington, DC: Author.

Heclo, H. (1977). *A government of strangers: Executive politics in Washington*. Washington, DC: Brookings Institution.

Joyce, P. G. (1990). An analysis of the factors affecting the employment tenure of federal political executives. *Administration and Society*, *22*, 127-145.

Lewis, D. 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, P. (1995). *Thickening government: Federal hierarchy and the diffusion of accountability*. Washington, DC: Brookings Institution.

Lucier, C., Schuyt, R., & Spiegel, E. (2003). Deliver or depart: CEO succession 2002. *Strategy+Business 31*, 32-45.

Mackenzie, G. C. (Ed.). (1987). *The in-and-outers: Presidential appointees and transient government in Washington*. Baltimore: Johns Hopkins University Press.

Maranto, R. (1998). Thinking the unthinkable in public administration: A case for spoils in the federal bureaucracy. *Administration and Society*, *29*, 623-642.

McCubbins, M. D., Noll, R., & Weingast, B. (1987). Administrative procedures as instruments of political control. *Journal of Law, Economics, and Organization*, *3*, 243-277.

McCubbins, M. D., & Schwartz, T. (1984). Congressional oversight overlooked: Police patrol versus fire alarm. *American Journal of Political Science*, *32*, 165-177.

Michaels, J. E. (1997). *The president's call: Executive leadership from FDR to George Bush*. Pittsburgh, PA: University of Pittsburgh Press.

Moe, T. M. (1982). Regulatory performance and presidential administration. *American Journal of Political Science*, *26*, 197-224.

Moe, T. M. (1985). The politicized presidency. In J. E. Chubb & P. E. Peterson (Eds.), *The new direction in American politics* (pp. 235-271). Washington, DC: Brookings Institution.

Moe, T. M. (1989). The politics of bureaucratic structure. In J. E. Chubb & P. E. Peterson (Eds.), *Can the government govern?* (pp. 267-329). Washington, DC: Brookings Institution.

Nathan, R. P. (1975). *The plot that failed: Nixon and the administrative presidency*. New York: John Wiley.

Nathan, R. 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 Institution.

Peacock, M. (2003, June 13). *Program performance and the FY 2004 budget process*. Forum presented at the Rayburn House Office Building, Washington, DC.

President's Foreign Intelligence Advisory Board. (1999, June). *Science at its best, security at its worst: A report on security problems at the U.S. Department of Energy.* Washington, DC: Author.

Pressman, J. L., & Wildavsky, A. (1974). *Implementation: How great expectations in Washington are dashed in Oakland.* Berkeley, CA: University of California Press.

Stanley, D. T., Mann, D. E., & Doig, J. W. (1967). *Men who govern: A biographical profile of federal political executives.* Washington, DC: Brookings Institution.

Tendler, J. (1997). *Good government in tropics.* Baltimore: Johns Hopkins University Press.

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

U.S. Office of Management and Budget. (2003a). *Budget of the United States government FY 2004: Performance management and assessments.* Washington, DC: U.S. Government Printing Office.

U.S. Office of Management and Budget. (2003b). *Budget of the United States government FY 2004: Rating the performance of federal programs.* Washington, DC: U.S. Government Printing Office.

U.S. President's Foreign Intelligence Advisory Board. (1999, June). *Science at its best, security at its worst: A report on security problems at the U.S. Department of Energy.* Washington, DC: Author.

U.S. Senate. (2000). *Policy and supporting positions.* Washington, DC: Author.

Wilson, J. Q. (1989). *Bureaucracy: What government agencies do and why they do it.* New York: Basic Books.

John B. Gilmour is an associate professor of government at the College of William & Mary. He is also the associate director of the Thomas Jefferson Program in Public Policy. He is the author of two books as well as articles published in American Journal of Political Science, Journal of Politics, Legislative Studies Quarterly, and Public Administration Review.

David E. Lewis is an assistant professor of politics and public affairs at Princeton University. His research interests include the presidency, executive branch politics, and public administration. He is the author of Presidents and the Politics of Agency Design (2003) and has published articles on American politics, public administration, and management in British Journal of Political Science, Journal of Politics, Political Analysis, Public Administration Review, and Enterprise and Society.