# Measuring the Impact of Appointee Vacancies on U.S. Federal Agency Performance[1]

In this paper we evaluate the relationship between vacancies in Senate confirmed positions (PAS) and U.S. federal agency performance. We explore this relationship using an innovative new perceptual measure of agency performance. This measure is derived from the evaluation of federal executives who work closely with the agencies they assess. The measure is comparable across agencies and avoids many of the limitations of existing measures. We find a robust correlation between vacancies in Senate-confirmed positions and lower evaluations of agency performance, even when accounting for differences in the way Republican and Democratic federal executives perceive performance. We conclude with a discussion of how dysfunction in the U.S. appointment process is influencing federal government performance on key tasks and the implications of our findings for the creation and use of performance measures in a world of partisan differences.

Mark D. Richardson, Georgetown University
mark.richardson@georgetown.edu

Christopher Piper, Partnership for Public Service
cpiper@ourpublicservice.org

David E. Lewis, Vanderbilt University
david.lewis@vanderbilt.edu

---

After sixteen months in office, President Biden had successfully filled 350 out of 799 key executive branch policymaking positions (Partnership for Public Service 2022). The U.S. Constitution requires that all principal officers in the executive branch such as cabinet secretaries and agency heads be nominated by the president and confirmed by the Senate. Prominent media outlets decried the slow pace of presidential nominations and Senate confirmations, with the *New York Times* asking, "Why are so many government positions still vacant?" and the *Washington Post* noting "A Senseless Logjam is Holding Up Nominations" (Bernstein 2022; Risen 2021). Outside observers raised concerns about how vacancies would create leadership gaps, implementation problems and poor agency performance (see, e.g., O'Connell 2021; Stier 2021). Others worried that the slow pace of appointments had adversely affected U.S. policy in areas as diverse as the conflict in Ukraine, the rollout of the infrastructure agenda, and economic policy (e.g., Gramer 2022; Hussein et al. 2022).

President Biden's pace of nominations was by no means an outlier. He was on par with his predecessors (Stier 2021). During the Trump Administration, for example, critics lambasted the president for the slow pace of executive appointments, but the president was largely unmoved, claiming he liked acting officials (Eilperin et al. 2019). Where Democrats saw vacancy-related performance problems, the Republican president saw a high performing administration even with vacancies. These struggles and concern for their consequences raises the more general question of how vacancies influence federal agency performance. This is an important question since vacancies—i.e., periods when a position lacks a Senate-confirmed appointee in a position requiring one—are a regular feature of modern administration and the consequences for performance can be quite broad if they systematically influence performance and policy (Dull and Roberts 2009; O'Connell 2009; Pfiffner 1996; Resh et al. 2021; Rutherford et al. 2018). Further, like the noted differences between Democrats and President Trump, the answer to whether and to what extent

vacancies impact agency performance is unresolved in the literature, with some scholars arguing that under certain conditions vacancies may improve performance. The temporary officials that fill in when there are vacancies may provide presidents additional flexibility and can provide extra competence or connections to the president in some circumstances (Kinane 2021; Mendelson 2014, 2020; Piper 2022).

Evaluating the impact of vacancies on performance has been difficult because appropriate measures of federal agency performance are elusive. Unlike private sector organizations, there is no profit equivalent that provides a shorthand way of comparing performance. This makes assessing the performance of government agencies with and without confirmed appointees challenging. The outputs of federal agencies are difficult to observe and measure across contexts (Wilson 1989). Scholars have made important progress measuring comparative performance through creative means but critics charge that such measures depend upon questionable self-reports, are limited to specific tasks or contexts that limit generalizability, or are hamstrung by partisan disagreements in defining good agency performance (Meier et al. 2015; Nyhan and Marlowe 1995; Boyne and Dahya 2002). It is no surprise that work systematically connecting vacancies to performance is rare (Rutherford et al. 2018; see, however, Piper and Lewis 2023; Wrightman et al. 2022).

In this paper we evaluate the relationship between vacancies in Senate confirmed positions (PAS) and federal agency performance using new perceptual measures of performance. These new measures overcome many of the limitations in existing approaches because they are not self-reports and are comparable across agencies, while still focusing performance on core agency missions. We rely on federal executives to evaluate the performance of agencies other than their own. Their perceptions of the performance of other agencies, given their own expertise and access, are informed by their own observations and experience. Their perceptions are important since common perceptions influence agency reputations, budgets, and politics (Bellodi 2022; Richardson et al.

2018). Perceptual measures have their own limitations, including partisan bias, something we explore in detail in this paper.

We find a robust correlation between vacancies in Senate-confirmed positions and lower evaluations of agency performance by federal executives familiar with agency activities. While there are some agencies where Republicans and Democrats disagree over performance, the effect of vacancies on performance remains substantial after accounting for these differences. We conclude by discussing how appointment process dysfunction is influencing federal government performance and how we should think about the creation and use of performance information in a world of partisan differences.

**Connecting Appointees and Vacancies to Performance**

The relationship between vacancies in Senate-confirmed positions and federal agency performance is an important one. Positions filled by appointees experience systematically higher turnover rates and longer vacancies than positions designated to be filled by career professionals (Heclo 1977; Lewis 2007; O'Connell 2009, 2020). Scholars are concerned that these vacancies make it more difficult to engage in long-term planning, generate lower morale, and reduce the incentives of outside stakeholders to invest time and resources in the agency (O'Connell 2009, 2020; Piper and Lewis 2023). In the absence of a Senate confirmed official in an executive position, temporary acting officials perform the duties of the position. So, executive positions are rarely vacant per se. More commonly, temporary replacements fill vacancies and debates about performance revolve around the value of a permanent Senate confirmed appointee versus temporary officials.[2]

---

[2] Under the Federal Vacancies Reform Act, vacant positions can be filled by acting officials for no more than 210 days. This time limit is extended during a presidential transition and the clock stops during a period when the president has a nomination pending in the Senate. Once the clock has

Other scholars, by contrast, have pointed out how vacancies or turnover can lead to improved agency performance in some contexts (Boyne and Dahya 2002; Boyne et al. 2011; Kinane 2021). If the appointed or career professionals that lead agencies during vacancies are more qualified or responsive than the appointees they replace, for example, this could help the agency (Mendelson 2014). Indeed, the career professionals that commonly serve as acting leaders while the president and the Senate negotiate over appointed leaders usually have long agency experience and significant policy and public management expertise (Lewis 2007). If appointed leaders unnecessarily meddle in agency processes and disrupt efficient routines, long vacant periods of depoliticization may improve the ability of the agency to tackle hard problems.

Efforts to evaluate these claims have been hindered by the difficulty of measuring performance across contexts. The outputs and outcomes of federal agencies are difficult to observe and compare (Wilson 1989). Some scholars have made progress by examining self-reported performance (see, e.g., Moynihan and Pandey 2005), but such efforts raise concerns about the connection between self-reports and actual performance (Meier et al. 2015). The proliferation of government-developed performance measures has provided another means of measuring performance (see, e.g., Boyne et al. 2011; Lewis 2007), but such measures can be subject to manipulation or politicization (see, e.g., Lavertu and Moynihan 2013). Others have focused on concepts measured in surveys of federal employees like the Federal Employee Viewpoint Survey (for a review see Fernandez et al. 2015). The measures have the virtue of being comparable across

---

expired, the authority of positions is delegated to a lower level agency official. In contrast to vacancies in executive agencies, there are no temporary officials filling vacancies on boards or commissions although some statutes allow members whose terms have expired to continue to serve for a discrete period or until a replacement is found.

agencies, but it is unclear how well such measures tap underlying performance. It is also not clear what organization respondents are evaluating when asked about their agency or organization (Thompson and Siciliano 2021).

Scholars have employed other creative measures of performance but these are often limited to one type of organization such as city services (Wrightman et al. 2022), law enforcement (Boylan 2004; Hur 2013) or schools (see, e.g., Meier and Hicklin 2007; Meier and O'Toole 2002; Rutherford 2016) or one type of task that may not be part of an agency's core mission such as budget forecasting (Krause and Douglas 2006), limiting payment errors (Park n.d.), or responding to FOIA requests (Wood and Lewis 2017). It is not clear whether results in such studies are applicable across agencies to the performance of core agency missions.

In a recent innovation, Bellodi (2022) introduces a way to measure a related concept -- agency reputation -- across time and contexts using the speeches of legislators. Agency reputation for Bellodi operates like affect—i.e., positive or negative sentiments about the agencies based upon what legislators say in floor speeches. Agency affect or reputation is conceptually distinct from performance since performance is only one component among many that comprise agency reputation. Differences in policy views and partisanship drive the content and sentiment in legislative speeches about agencies like the Environmental Protection Agency and or the Consumer Financial Protection Bureau. We can predict what Republicans and Democrats will say about these agencies without knowing anything specific about these agencies' performance. Predictably, the opposition party becomes more positive about agencies once their party occupies the White House. What legislators say about agencies and resulting measures of reputation built on these statements is something quite different than performance.

What is needed is an evaluation of the impact of vacancies on performance, using properly validated measures and analysis that is robust to partisan differences in evaluation.

**How do Vacancies Influence Performance?**

In June 2021, the Senate confirmed Kiran Ahuja, President Biden's nominee to be director of the Office of Personnel Management (OPM), in a party-line vote (Wagner 2021). During the period from January to June, OPM was led by a temporary acting official, Kathleen McGettigan. McGettigan was a career executive with long experience in the agency. OPM was fortunate since the president nominated and the Senate confirmed Ahuja quickly relative to other executive positions. Many other executive branch positions were still without a Senate confirmed leader 18 months into Biden's presidency. When there are vacancies in executive agencies, temporary officials like McGettigan step in and fill roles under strict rules for length of service, limits that can be extended if a vacancy occurs at the start of a term or if the president has made a nomination. Whereas Senate confirmed appointees usually serve at the pleasure of the president (as long as the president allows), the Vacancies Act sets limits on the tenure of acting officials. McGettigan took on the role of acting director knowing her tenure would be short, serving only until a replacement was confirmed.[3]

The case of McGettigan raises two issues central to this paper. First, does the presence of Ahuja, a permanent Senate-confirmed appointee, matter for OPM's performance? Second, how do we connect the presence or absence of a Senate confirmed leader to performance when Republicans and Democrats define good performance in OPM differently? In this section, we address these

---

[3] Some acting officials occasionally serve as long or longer than some Senate-confirmed appointees (Kinane 2021, O'Connell 2020). As Kinane (2021, 601) notes, "an interim can serve for almost two years or even indefinitely…" while the average appointee serves 2-3 years (Dull et al. 2012). Acting officials, while occasionally serving long tenures, average much shorter tenures than appointees (O'Connell 2020) and others perceive them to be short timers because they are always operating under a time limit or with a pending nomination over their head.

questions. We explain why vacancies can be harmful for performance and how this should be true even accounting for differences in how Republicans and Democrats define good performance. We summarize our expectations in three hypotheses.

There are several ways periods of temporary leadership can influence agency performance. First, while some temporary leaders can be political appointees chosen strategically by the president (Kinane 2021; Piper 2022), acting officials generally know their tenure will be short (O'Connell 2020). This systematically shapes their time horizons and what they attempt to accomplish during their tenures (Piper and Lewis 2023). For example, when McGettigan assumed control of the OPM in January 2021, the OPM had a longstanding problem with backlogs in federal retirement claims (Fahrenthold 2014; Weisner 2023). These backlogs stemmed from an antiquated system that required processing all claims by hand in an abandoned limestone mine 230 feet below the ground in Boyers, PA. This is one of the biggest management challenges facing the agency. It also cannot be solved in 6 months or 12 months. Automating the process is a complex job, requiring the cooperation of hundreds of agencies and an intricate procurement and implementation process. Given McGettigan's perspective, this is not a problem she can tackle directly since she cannot possibly see the reform through. Many management challenges, from difficulties in human resources to information technology systems to financial management, cannot be solved in a year.

Expected short tenures also influence the way that career employees operate. Had McGettigan launched a major initiative to reform the retirement claims process, there was a reasonable chance that the agency might change course once the Senate confirmed a new leader. Career executives understand this. The new confirmed leader would naturally have her own priorities, some influenced by the confirmation process. McGettigan's successor might not prioritize retirement claims, preferring instead to focus on data security or civil service reform. In addition, if Ahuja was to prioritize reforming the system, she might opt for a different solution, perhaps

increasing employment rather than tackling the difficult information technology problem. This uncertainty makes the continuing professionals in government cautious. These professionals do not want to waste effort on initiatives that will be abandoned by the confirmed leader. They also recognize that temporary leaders cannot credibly commit to either rewarding or punishing them for hard work or slow walking the acting director's initiatives.

The combined effect of short tenure on the acting leader and agency employees often leads acting leaders to focus on keeping current programs running, holding the agency in place until a replacement is confirmed. Like a temporary teacher or interim dean, acting leaders are reluctant to take on major issues, preferring instead to maintain existing programs and activities. This can be beneficial because acting leaders like McGettigan are often quite experienced and effective. It does mean, however, that agencies delay solving hard and slow moving problems (Piper and Lewis 2023; for a few counter examples, see O'Connell 2020).

The health of the agency and its performance will depend upon the support of outside stakeholders. For example, any effort to reform the retirement claims system will depend upon support from Congress, federal employee unions, area non-profits, and other agencies. Securing cooperation from these parties takes time and effort. It also takes credibility. These groups are inveterate watchers of Washington. They, too, understand that someone like McGettigan will only likely be in office for a short period. They would prefer not to invest time in a reform effort until they know that their expenditure will pay off. Members of Congress will be slow to approve increased appropriations or changes to Title 5 without a full plan and a commitment to seeing it through. Non-governmental groups will not want to make costly public commitments to policies that may not come to fruition. No agencies, other than OPM, prioritize federal retirement claims but they all play a role in the retirement process, collecting and sending materials to OPM. They all will

have to change their processes in any reform. Securing their cooperation will be difficult, particularly when an agency has no confirmed leader (O'Connell 2020; Piper and Lewis 2023).

OPM was fortunate to have a presidential nominee early in President Biden's term and relatively quick Senate confirmation. Other agencies are not so fortunate. Long delays in nominations and confirmations send a signal to federal employees, particularly those in decision making roles, about how much priority elected officials place on the work that they do. Agencies regularly neglected in this process make inferences about the value of their work. For employees motivated partly by public service and policy, this can operate like a cut in compensation. Of course, career acting officials may be well liked and for executives moving into top jobs in an acting capacity, this might improve morale and be an inducement to stay (O'Connell 2020; Rutherford and Hameduddin 2022). On average, however, long vacancies are correlated with lower morale (Piper and Lewis 2023).

Given the prevalence of vacancies, presidents are increasingly turning to the use of political appointees to serve as acting officials and in other agency positions (Kumar 2021; Moore 2018). These other appointees are responsive to the president since they serve at the president's pleasure and have no competing obligations to the Senate (Kinane 2021; Mendelson 2020; Piper 2022). They do, however, confront the same short time horizons of other acting officials and often have fewer qualifications than their careerist counterparts (Mendelson 2020). Appointees in staff positions can help presidents accomplish their policy goals but politicization itself is correlated with poor agency performance, perhaps because these appointees focus on the president's political and policy priorities rather than agency performance (Lewis 2007).

In total, and while there are exceptions, vacancies generally hurt agency performance through short time horizons, less outside investment, and lower morale.

H1: Federal agencies with long vacancies in Senate confirmed leaders will have lower performance than other agencies.

Of course, this raises the interesting question of whether agencies regularly operating without the expectation of appointed leadership adapt to a new normal. Do leaders and employees take a longer view and outside stakeholders adjust as well?

*Can Republicans and Democrats Agree on Good Performance?*

A natural question, given Republican objections to Ahuja's nominations and her plans for the federal workforce, is whether some Republicans prefer an OPM that does less. For some, an OPM under McGettigan may be a better performing OPM than one under Ahuja. This makes evaluating the impact of vacancies on performance difficult. The case of OPM, however, helps explain why it is possible to measure performance even given policy disagreement among stakeholders. While Republicans and Democrats may disagree about Biden Administration policies toward diversity, unions or federal employee rights, they agree about a lot of what OPM and other agencies do. For example, both parties prefer that OPM process retirement claims quickly and accurately. They prefer that OPM protect the data of federal employees. They may disagree about the meaning of Title 5 but agree that agencies should implement the law.

A large percentage of government work has little to do with disagreements about law and policy. It is apolitical and includes tasks like procuring goods, managing programs, coordinating systems. As a general matter, elected officials prefer that agencies do a good job at those things. They want agencies to have effective human resources systems, buy goods and services cheaply, and use technology effectively. In addition, most federal government programs are overwhelmingly popular with the public (Gramlich 2017; Light 2008), suggesting that for most programs partisans agree on good performance when they see it. Every federal program was at one time supported by majorities in the House and Senate and the president. The public and most members of Congress are generally supportive of federal efforts to approve patents, protect ports, build infrastructure, and provide clean and safe natural parks. While elected officials have an incentive to focus on areas of

disagreement, this should not overshadow the vast number of things about which they agree. Indeed, bipartisan groups of members request hundreds of GAO investigations each year, showing joint concern for the efficacy of federal programs. If we separate out Republicans and Democrats evaluating performance, they should agree a majority of the time on what they observe and draw the same conclusions about the impact of vacancies on performance.

> H2: Republicans and Democrats generally agree on which agencies are performing well or poorly.

> H3: The negative relationship between vacancies and performance is robust even when accounting for the effects of partisanship on performance evaluation.

**Data, Variables, and Methods**

To overcome existing measurement problems and effectively evaluate the relationship between vacancies in PAS positions and performance, we introduce a new measure of federal agency performance relying on federal executive experience. We begin by describing how we generated this measure. We regress this new perceptual measure of agency performance on data on vacancy lengths. We then evaluate whether this measure is influenced by the partisanship of those evaluating federal agency performance.

*Measuring Federal Agency Performance*

In 2020, in collaboration with the Partnership for Public Service and academic colleagues, we fielded a survey targeting all political appointees and senior career managers running agencies, offices, and programs in the executive branch. The response rate for the survey was 9.1% (1,485 completed surveys out of 16,232) and the participation rate (i.e., the percentage that completed at

least part of the survey) was 11.5% (1,861 complete or partial surveys out of 16,232), rates

comparable to most public opinion telephone surveys (AAPOR 2017).[4]

Federal executives working at the highest level of the executive establishment are uniquely

situated to observe agency performance, perhaps better than any other population. On average, they

have X years of experience working in their agencies and Y years in the federal government overall.

They interact with officials in other agencies every day, on Zoom calls, in meetings and cooperate

with them on interagency projects and joint rulemakings, negotiating over shared jurisdiction and

responsibility. Their own experience working with other agencies informs their views. Of course,

factors other than objective performance such as partisanship can also shape perceptions and some

executives will be more knowledgeable than others. So, while utilizing the expertise of

knowledgeable insiders presents a great opportunity, we must do so in a way that accounts for

variation in the quality of executive opinions, pays special attention to validation, and addresses the

thorny issue of partisanship in evaluations.

To begin, we tried to narrowly target the expertise of federal executives. We asked our high-

level executives to identify the agencies they worked with the most. Specifically, the survey asked:

"Please select the three agencies you have worked with the most in order of how often you work

---

[4] One potential concern with the survey was that it was in the field during the pandemic. This might

have influenced respondent perceptions of agency performance. While the pandemic may have

increased or decreased perceptions of performance overall, this would only be a problem for

inference if COVID had *differential* effects on perceptions of performance that are correlated with

vacancies. In other words, COVID can artificially increase or decrease overall perceptions, but the

estimates of vacancies on performance can still be unbiased.

with them."[5] Later in the survey, after asking federal executives about the performance of their own agencies, we then asked them to evaluate the performance of the agencies they had mentioned at the start of the survey, plus two others. Specifically, they were asked, "How would you rate the overall performance of the following agencies in carrying out their missions?" and given options from 1- Not at all effective to 5-Very effective. They were also provided a Don't know option. Each respondent rated up to 5 agencies.

We use these ratings to generate numerical estimates of agency performance, adjusting for differences in the ability of federal executives to rate performance and differences in the way that federal executives use the 1 to 5 scale. Indeed, some federal executives may not perceive performance clearly but rate agencies anyway. Others may be harder or easier graders (e.g., what level of performance is necessary to be rated a 1, 2, 3, 4 or 5?). We have ratings on 179 agencies from 1,379 raters and 4,555 ratings of federal agency performance overall. We only report data for agencies with at least 5 performance ratings.

We use these ratings to estimate agency performance using a Bayesian multi-rater item response model. We leave most of the details for Appendix A but note that we estimate a hierarchical model to account for the fact that raters share a workplace and we use informed priors on latent performance to give additional weight to those respondents working closely with the agency being evaluated (as compared to those that evaluated an agency selected at random).[6]

---

[5] For full details of these "works with" question and rating question, see Appendix A.

[6] In Appendix C we compare hierarchical and non-hierarchical models as well as estimates with informed and naïve priors. The correlations among the estimates from these different models range from 0.97 to 0.99 and so make little substantive difference in the estimates.

In Figure 1 we include the resulting numerical estimates of federal agency performance (Mean 0.01; SD 0.78; Min -1.96; Max 1.41) and in Appendix B we include the numerical estimates themselves. In the figure the dots represent the estimates and the bars represent the degree of uncertainty about the estimate. The uncertainty can derive from there being few ratings or because there is disagreement among the raters, or both. Indeed, it is possible, for example that Republicans and Democrats rate performance differently. This is something we address in more detail below.

Among the highest performing agencies, according to federal executives, are the National Security Agency, the National Cemetery Administration (VA), and the National Institutes of Health (NIH). The top-10 also includes several technical, science, and military agencies. Among the lowest performing agencies are a part of the Executive Office of the President whose chair was vacant for a significant part of the Trump Presidency (CEA), the agency responsible for environmental cleanup (OEM) in the nation's nuclear weapons labs (whose PAS position was vacant at the time of the survey), and an agency that has had 9 PAS heads since 2015 (OPM). There is at least *prima facie* evidence that there is a relationship between PAS vacancies and federal agency performance.

**Figure 1. Expert Opinion Estimates of 2020 Federal Agency Performance**

## Latent Agency Performance

### Latent Performance



(Left panel, axis −3 −2 −1 0 1 2 3)

National Security Agency
National Cemetery Administration
National Institutes of Health
Bureau of Labor Statistics
Office of Legal Counsel
Coast Guard
National Science Foundation
Office of Science
Executive Office for United States Attorneys
National Aeronautics and Space Administration
Bureau of Economic Analysis
National Agricultural Statistics Service
Office of Energy Efficiency and Renewable Energy
U.S. Patent and Trademark Office
Federal Deposit Insurance Corporation
Navy
Federal Reserve
Government National Mortgage Association
Defense Advanced Research Projects Agency
Bureau of Consular Affairs
Natural Resources Conservation Service
Air Force
Central Intelligence Agency
Defense Logistics Agency
Defense Contract Management Agency
Federal Aviation Administration
Federal Trade Commission
National Institute of Standards and Technology
Food and Drug Administration
Army
National Park Service
National Oceanic and Atmospheric Administration
Secret Service
Animal and Plant Health Inspection Service
Department of Defense
Bureau of Reclamation
Office of Special Counsel
Millennium Challenge Corporation
Mine Safety and Health Administration
Office of the Comptroller of the Currency
U.S. Geological Survey
Nuclear Regulatory Commission
Institute of Education Sciences
Agricultural Research Service
Political Affairs
National Nuclear Security Administration
U.S. Fish and Wildlife Service
Bureau of Ocean Energy Management
Advanced Research Projects Agency-Energy
Forest Service
Centers for Medicare and Medicaid Services
Office of Electricity
Employment and Training Administration
Bureau of Diplomatic Security
Joint Chiefs of Staff
Food Safety and Inspection Service
International Trade Administration
Administration for Children and Families
U.S. Census Bureau
Health Resources and Services Administration
Defense Contract Audit Agency
Federal Bureau of Investigation
Federal Highway Administration
Management
Bureau of the Fiscal Service
Department of the Treasury
Defense Commissary Agency
Bureau of International Narcotics and Law Enforcement Affairs
Office of Government Ethics
Department of Energy
Office of the Director of National Intelligence
Economic Research Service
Public Diplomacy and Public Affairs
Veterans Health Administration
Department of State

### Latent Performance (bottom axis −3 −2 −1 0 1 2 3)

## Latent Agency Performance

### Latent Performance

(Right panel, axis −3 −2 −1 0 1 2 3)

Department of Energy
Office of the Director of National Intelligence
Economic Research Service
Public Diplomacy and Public Affairs
Veterans Health Administration
Department of State
Defense Finance and Accounting Service
Small Business Administration
National Highway Traffic Safety Administration
United States Agency for International Development
Internal Revenue Service
Federal Transit Administration
Office of the Secretary of Defense
United States Postal Service
Defense Health Agency
Department of Agriculture
Office of the United States Trade Representative
Agricultural Marketing Service
Citizenship and Immigration Services
Federal Housing Administration/Office of Housing
Department of Transportation
Farm Service Agency
Department of Labor
Centers for Disease Control and Prevention
Department of Veterans Affairs
Federal Emergency Management Agency
Economic Growth, Energy, and the Environment
National Endowment for the Humanities
Department of Justice
Occupational Safety and Health Administration
Customs and Border Protection
Drug Enforcement Administration
National Archives and Records Administration
Food and Nutrition Service
Department of the Interior
Council on Environmental Quality
Social Security Administration
Federal Motor Carrier Safety Administration
Equal Employment Opportunity Commission
Department of Health and Human Services
Office of Federal Student Aid
Civilian Security, Democracy, and Human Rights
Arms Control and International Security Affairs
Consumer Financial Protection Bureau
Department of Commerce
Economic Development Administration
Transportation Security Administration
Department of Homeland Security
Federal Labor Relations Authority
Office of Workers' Compensation Programs
Rural Development
Cybersecurity and Infrastructure Security Agency
Bureau of Land Management
Office of Elementary and Secondary Education
Office of Management and Budget
National Security Staff
Immigration and Customs Enforcement
Indian Health Service
Veterans Benefits Administration
Merit Systems Protection Board
Office of National Drug Control Policy
Office of Science and Technology Policy
Department of Education
Department of Housing and Urban Development
Wage and Hour Division
Office of Nuclear Energy
Bureau of Indian Affairs
Environmental Protection Agency
General Services Administration
National Transportation Safety Board
Council of Economic Advisers
Office of Environmental Management
Office of Personnel Management

### Latent Performance (bottom axis −3 −2 −1 0 1 2 3)

Note: Numerical estimates of federal agency performance based upon 4,310 ratings by 1,239 federal executives in response to the question: "How would you rate the overall performance of the following agencies in carrying out their missions?" and given options from 1-Not at all effective to 5-Very effective. Federal executives were also provided a Don't know option. Ratings were estimated using a Bayesian hierarchical multi-rater item response model with informed priors (based upon those with specific expertise about an agency).

To validate these measures, we compare them to three other measures of agency health and performance (Figure 2). First, we compare the performance ratings to 2020 agency average self-reported performance. The survey asked respondents, "How would you rate the overall
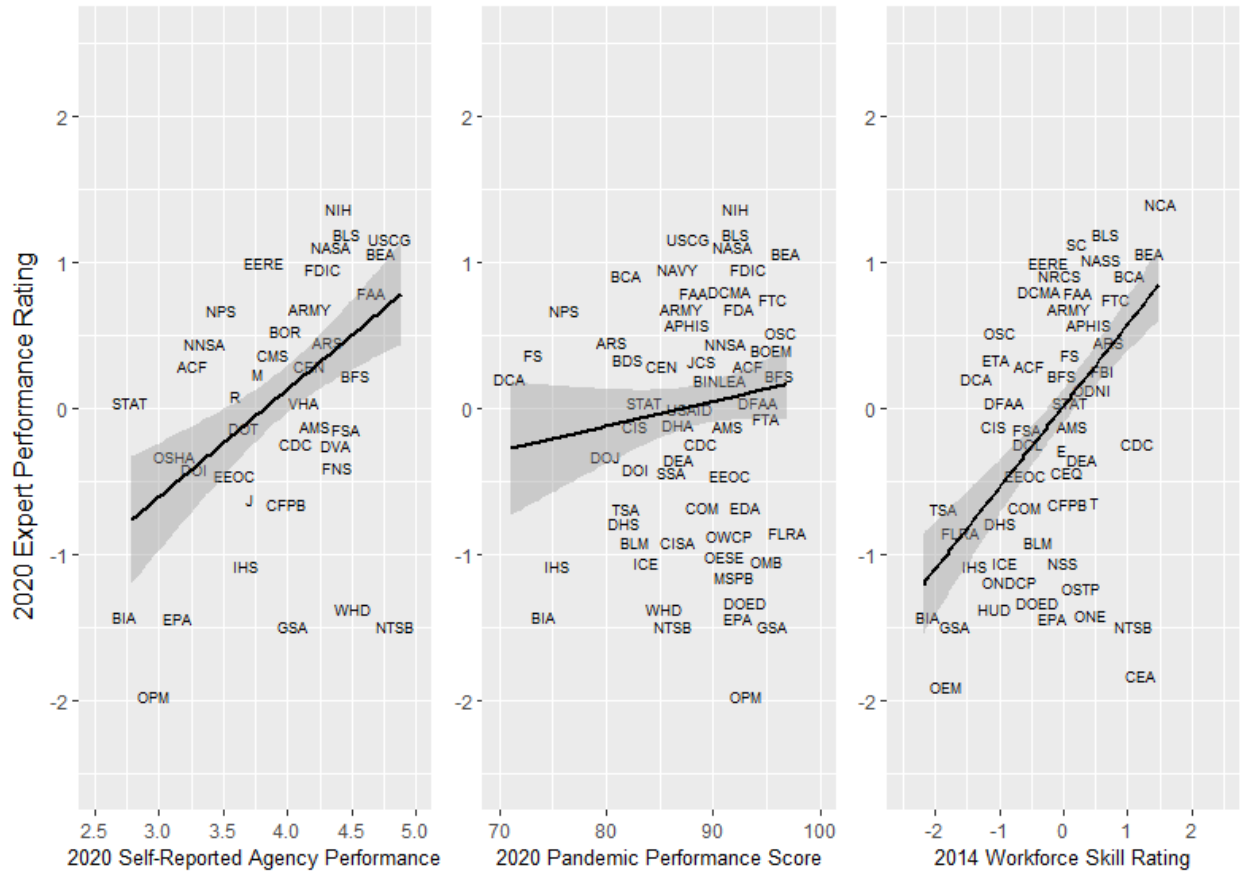
performance of [your agency] in carrying out its mission?" The question asks federal executives to

evaluate how well their agencies are doing by name (i.e., [your agency] is replaced with the actual

name of the respondent's workplace). Limiting our sample to those agencies with at least 5

respondents leaves us with 91 agencies. Second, we correlate the new measure with COVID

performance as evaluated by the Partnership for Public Service. In 2020, the Partnership for Public

Service produced scores for agency COVID performance (Partnership for Public Service 2020). The

Partnership's scores are an index created using federal employee responses to the following

questions:

- During the COVID-19 pandemic, my work unit has met the needs of our customers.
- During the COVID-19 pandemic, my work unit has contributed positively to my agency's performance.
- During the COVID-19 pandemic, my work unit has produced high-quality work.
- During the COVID-19 pandemic, my work unit has adapted to changing priorities.
- During the COVID-19 pandemic, my work unit has successfully collaborated.
- During the COVID-19 pandemic, my work unit has achieved our goals.

The overall score reflects the views of the rank-in-file about how they did during the difficult 2020

pandemic period. Finally, we compare estimates of performance to estimates of workforce skill at

the end of the Obama Administration. In 2018, Richardson et al. used a survey of federal executives

to ask, "In your view, how skilled are the workforces of the following agencies?" and given options

from 1-Not at all skilled to 5-Very skilled. They also provided a Don't know option. Each

respondent rated up to 8 agencies, providing thousands of ratings of different agencies. Richardson

et al. (2018) used these ratings to develop numerical estimates of workforce skill. We correlate these

estimates of workforce skill at the start of the Trump Administration with our new measures of

performance.[7]

**Figure 2. Comparison of Expert Estimates of Agency Performance vs. 2020 Self-reported Performance, 2020 Best Places to Work Pandemic Performance Scores, and 2014 Workforce Skills Ratings**



Note: y-axis is numerical estimate of agency performance based upon executive responses to the question: "How would you rate the overall performance of the following agencies in carrying out their missions?" and given options from 1-Not at all effective to 5-Very effective. Federal executives were also provided a Don't know option. The x-axis of three graphs are 1) 2020 average responses to question: "How would you rate the overall performance of [your agency] in carrying out its mission?"; 2) a 2020 pandemic performance score from Partnership for Public Service; and 3) a numerical estimate of workforce skill based upon 2014 federal executive responses to question: "In your view, how skilled are the workforces of the following agencies?" (Richardson et al. 2018). Fitted lines for self-reports estimated based upon all agencies with at least 5 respondents in the survey. Some agency acronyms are excluded from the figure to limit overlap among acronyms.

---

[7] We have also compared the measure to the Government Accountability Office's (GAO) data on

open recommendations in 2020. Those results confirm what is reported here and are included in

Appendix I.

The new expert performance ratings are positively correlated with all three measures, though strongest with federal executive self-reports and previous expert ratings for workforce skill. Generally, when executives rated their own agencies poorly, outsiders agreed. When federal executives rated their agencies well, so did outsiders. While there are some outliers where agencies rated themselves as performing well but experts did not (e.g., National Transportation Safety Board [NTSB] and Wage and Hour Division [WHD]) and vice versa, the strong correlation suggests that people inside and outside agencies are observing the same performance (corr 0.46). Reassuringly, expert evaluations of the quality of the workforce before the start of the Trump Administration are also correlated positively with 2020 performance (corr 0.67). Notable differences between workforce skills in the Obama Administration and performance in the Trump Administration include agencies in the Executive Office of the President like the Council of Economic Advisers (CEA), National Security Staff (NSS), and Office of Science and Technology Policy (OSTP). Other notable differences include the Centers for Disease Control (CDC) and Prevention and the Environmental Protection Agency (EPA), suggesting these agencies were less successful performing their core missions than their workforce capacity might have predicted at the end of the Obama Administration. The correlation between overall agency evaluations of COVID responses are also positively correlated with overall performance, but only weakly so (corr 0.18). This is not entirely surprising since we survey executives and the Partnership's score is based upon surveys of the entire workforce. We also focus on performance of an agency's core mission and the Partnership is focused on COVID performance. On their face, then, these measures of agency performance appear plausibly related to actual performance.

*Appointee Vacancies*

To connect survey responses to vacancies, we collected data on vacancies in the Trump Administration. We used the 2016 congressional publication *Policy and Supporting Positions* (i.e., Plum

Book) to identify the closest PAS appointee to every agency. We systematically tracked the occupants of each position (including acting officials) from January 20, 2017 until the soft launch of the survey on June 22, 2020. We counted the number of days the position lacked a Senate-confirmed head. For simplicity, we divided the number of days by 30 to report vacancy length in months. For commissions, we focused on length of time the chair position was vacant.[8] Out of a maximum of 42 months, the average position was vacant 18 months (SD 13.29), confirming the regularity with which agencies experience vacancies in leadership positions. The raw correlation between vacancies and performance is -0.11, suggesting a modest negative effect on performance. We estimate models that also evaluate the linearity of this relationship since it is possible that vacant months become less or more important as time goes on.

One immediate concern is that poor performance might *cause* vacancies rather than be caused by vacancies. For example, President Trump might have had a more difficult time finding persons to fill leadership positions in agencies that were struggling. To evaluate this possibility, we regressed Trump Administration vacancies on Obama Administration workforce skills ratings and other covariates to see whether agencies with lower human capital and, therefore, likely poor performance during the Obama Administration caused Trump Administration vacancies. While the estimated coefficient is negative, suggesting that high performing agencies had shorter vacancies, we

---

[8] We have also estimated models omitting the independent commissions and the estimates are substantively similar.

cannot reject the null that struggling agencies during the Obama Administration were no more likely to experience vacancies in the Trump Administration.[9]

*Controls*

To hone in on the effect of vacancies themselves we control for several features of the affected positions. First, we include a control for whether the nearest appointee vacancy is in the agency itself (0,1; 78%) or a higher level. For example, in our data the Bureau of Ocean Energy Management (BOEM) is headed by a career professional and the closest presidential appointee is the Assistant Secretary for Land and Minerals Management. The Bureau of Land Management (BLM), however, is overseen by the same assistant secretary but the BLM is headed by its own Senate-confirmed appointee. Second, to parse out the unique effect of a vacancy from turnover we control for the turnover in the position, which is the total number of persons that served in that position, acting or confirmed, during the Trump Administration from Inauguration Day until the start of the survey (Mean 1.53; SD 1.1; Min 0, Max 4).[10]

There are a number of agency characteristics that might be correlated with both vacancy length and agency performance. We include controls for several pre-treatment covariates, including agency structure, whether the agency implements a policy that President Trump mentioned during the 2016 campaign, agency workforce skills measured during the Obama Administration, employment and agency ideology. We include controls for whether the agency is an agency in the

---

[9] Despite the null finding, we also conduct mediation analysis to estimate how large the effect of this potential endogeneity might be relative to the direct correlation between vacancy length and performance. [Mark]

[10] For details of data collection on vacancies and turnover see Appendix F. We have also estimated models accounting for vacancies at different levels and include that analysis in Appendix G.

Executive Office of the President (0.05), whether the agency rated is an entire department (e.g., "Department of Agriculture (All)"; 0.11), or in an independent commission (0.08). The inclusion of the structural features means the base category is a sub-component of an executive department or an independent executive agency like the Environmental Protection Agency. Since the president's decisions about whether to fill vacancies and support the agency more generally could influence performance, we include indicators for whether the department (0,1; 0.65) implemented a policy that was a priority of the president during the 2016 election.[11] We include the 2014 workforce skills rating mentioned above as a measure of the health of the agency at the end of the Obama Administration. Finally, since the persistence of vacancies and presidential attention might be influenced by the ideological contours of what agencies do, we include controls for agency ideology measured prior to the start of the Trump Administration (Richardson et al. 2018). Richardson et al. asked executives during the Obama Administration about the ideological leanings of agencies they work with, and whether the "lean liberal, lean conservative, or neither consistently across Democratic and Republican administrations." They aggregate responses with a method similar to the workforce skill scores described here to generate estimates of agency ideology.

*Methods*

To evaluate the relationship between vacancies and performance, we estimate a series of models with OLS. For each hypothesized relationship, we estimate models on the complete set of agencies and the subset of agencies directly led by an appointee. We cluster the standard errors to account for the fact that different agencies are not completely independent since many are

---

[11] We identify policy priorities of the president using the Trump Campaign's 2016 Contract with the American Voter (https://assets.donaldjtrump.com/_landings/contract/O-TRU-102316-Contractv02.pdf).

subcomponents of larger departments. Specifically, we have 17 clusters, one for each executive department, one for the EOP, and one for independent agencies.

**Results**

We include the initial set of model estimates in Table 1. To begin, there are several interesting results among the controls. First, agencies directly led by appointees are estimated to have lower overall performance ratings, about 0.17 (~1/5$^{th}$ a standard deviation), although the estimates are imprecise. The estimates suggest that increasing penetration of appointees is correlated with lower perceptions of agency performance. Second, agencies that had higher leader turnover were estimated to have *higher* performance rating. Each additional leader during this 42-month period was correlated with a 0.10 higher performance. While scholars generally associate turnover with lower performance, turnover in this case could be a measure of whether agencies have presidential attention at all since such a large number of positions were vacant. The dynamics of turnover and the dynamics of vacancies might be quite different since long vacancies can be correlated with low turnover and vice versa (Piper and Lewis 2023; Rutherford et al. 2018). It is also possible that we miss some turnover because the data are quarterly. If the poorest performers also have the highest turnover, this could lead us to underestimate the negative effects of turnover on performance.

Also notable, model estimates reveal that agencies in the EOP, larger departments, and independent commissions had lower performance ratings than other agencies. Agencies in President Trump's EOP were estimated to have three quarters of a point lower performance ratings than bureaus in executive departments or independent administrations (i.e., the base category). Whole departments, when evaluated, were rated between 0.54 to 0.68 lower than other agencies and independent commissions, about 0.18 lower. In addition, agencies that were responsible for implementing policies the president mentioned during the campaign were estimated to have lower

performance, about half to a third of a point lower. Not surprisingly, agencies evaluated to have

skilled workforces by executives at the end of the Obama Administration were also rated to have

higher performance by executives during the Trump Administration. A standard deviation increase

in workforce skill (0.76) is estimated to increase performance by 0.42. Finally, agency ideology is

largely uncorrelated with performance.

**Table 1: OLS Models of the Effect of Vacancies on Agency Performance**

| Agency Performance (-1.96, 1.41) | All Agencies | All Agencies | PAS Head Only | PAS Head Only |
|---|---|---|---|---|
| Months Vacant (0-42) | -0.012 | -0.029 | -0.014 | -0.037 |
| | [0.003]** | [0.010]** | [0.005]** | [0.012]** |
| Months Vacant^2 | | 0.0004 | | 0.0005 |
| | | [0.0002]* | | [0.0002]** |
| *Position Controls* | | | | |
| Direct PAS Leader (0,1) | -0.177 | -0.185 | | |
| | [0.122] | [0.116] | | |
| Leadership Transitions (0-4) | 0.09 | 0.105 | 0.104 | 0.124 |
| | [0.034]** | [0.035]** | [0.041]** | [0.042]** |
| *Agency Controls* | | | | |
| EOP (0,1) | -0.779 | -0.736 | -0.765 | -0.713 |
| | [0.088]** | [0.082]** | [0.094]** | [0.082]** |
| Cabinet Department (0,1) | 0.356 | 0.367 | 0.379 | 0.385 |
| | [0.067]** | [0.058]** | [0.075]** | [0.064]** |
| Whole Department (0,1) | -0.54 | -0.615 | -0.586 | -0.675 |
| | [0.132]** | [0.136]** | [0.149]** | [0.158]** |
| Independent Commission (0,1) | -0.177 | -0.181 | -0.195 | -0.203 |
| | [0.074]** | [0.070]** | [0.078]** | [0.073]** |
| Priority Department (0,1) | -0.31 | -0.309 | -0.31 | -0.306 |
| | [0.137]** | [0.128]** | [0.151]* | [0.139]** |
| Workforce Skill -- Obama Administration | 0.551 | 0.54 | 0.583 | 0.572 |
| | [0.063]** | [0.063]** | [0.068]** | [0.068]** |
| Agency Ideology (L-C) | 0.045 | 0.038 | 0.008 | -0.004 |
| | [0.073] | [0.074] | [0.089] | [0.091] |
| Constant | 0.309 | 0.406 | 0.141 | 0.253 |
| | [0.152]* | [0.131]** | [0.064]** | [0.052]** |
| $R^2$ | 0.49 | 0.49 | 0.48 | 0.49 |
| N | 125 | 125 | 101 | 101 |

Note: * $p<0.1$; ** $p<0.05$. Dependent variable is agency performance estimate based upon aggregated responses to question: "How would you rate the overall performance of the following agencies in carrying out their missions?" (1-5). Models

estimated with Ordinary Least Squares and standard errors are clustered by department. Models 3, 4 estimated using only agencies whose immediate head is a Senate-confirmed political appointee (PAS).

Most importantly for the purposes of the paper, however, the models reveal a robust relationship between vacancy length and perceptions of agency performance. The coefficient estimate for the direct effect of the number of vacant months is negative and statistically significant in all four models. The result is strongest when subsetting to cases where the agency is led directly by a political appointee (Models 3, 4). Collectively, these results indicate that experts rate agency performance more poorly where PAS positions are vacant for long periods. To contextualize the effect, it is useful to evaluate the effect in terms of real vacancy times. For every 12 months a PAS headed position is vacant, the agency's average performance is estimated to decrease by between 0.14 and 0.17 on a scale of -1.96 to 1.41. If the position stays vacant for 24 months or about half the president's term (i.e., about 33% of the data), the estimated effect is between 0.28 and 0.34, or about one third of a standard deviation decrease. Finally, if the position is vacant for the duration of the Trump administration prior to the survey (i.e., about 13% of the data), the estimated performance is about 0.50 lower than other agencies, or about 2/3 of a standard deviation. This is important evidence, using new performance measures, that vacancies are correlated with lower federal agency performance. This confirms some of the theoretical expectations of past work (O'Connell 2009, 2020; Rutherford et al. 2018).

**Figure 3. Estimated Impact of Vacancies in Senate Confirmed Positions on 2020 Agency Performance Rating**
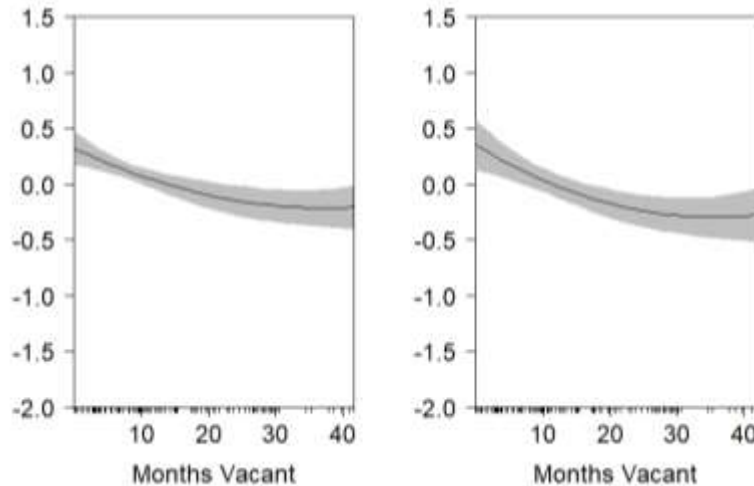
Note: Figure graphs predicted agency performance rating by months most appointed position vacant during the Trump Administration. The first model is based upon estimates from Model 2 in Table 1 and the second model is based upon estimates from Model 4 in Table 1.

While the average effect of vacancies is notably negative, the effect is non-linear (Models 2, 4; Figure 3). The effect is largest early in the vacancy and becomes negligible the longer a position is vacant. While we do not observe vacancies that carry over beyond the Trump Administration, there is suggestive evidence that the effect of a vacancy may turn positive if a position is vacant long enough. This may indicate that positions with persistent vacancies perform better if leadership positions were converted to positions filled by permanent career professionals.

One shortcoming of focusing on vacancy length overall is that it can conflate the timing of a vacancy with the length of a vacancy. If a position has been vacant 6 months, for example, this vacancy could be at any point in the Trump Administration up to the time of the survey. However, if a position was vacant 40 months, this implied a vacancy at the start, middle, and end of the term. We have done some additional analysis in Appendix H disentangling the timing of vacancies. It shows that vacancies at the start and vacancies at the end both have consequences for performance but, the vacancies later in the term mattered more for performance. The curvilinear relationship between vacancies and performance in Figure 3 exists for both types of vacancies.

One potential interpretation of these results is that our respondents see persistent vacancies and infer poor performance rather than actually observe poor performance. While this is possible, we think it is unlikely to explain all of this result for a couple of reasons. First, our measure of perceived performance is correlated with anecdotal accounts of poor performance in agencies with and without vacancies (e.g., Office of Personnel Management, Bureau of Indian Affairs) and other measures of performance less plausibly related to perceptions about vacancies. For example, our measure of performance is correlated with managers' own reports on performance. Self-reports are unlikely to be caused managers' inferences about vacancies. Second, if there is bias it may operate in the other direction. That is, federal executives may attribute *higher* performance to agencies led by experienced career professionals, leading us to underestimate the true size of the correlation.

In total, the evidence suggests a negative relationship between an agency's leadership vacancy and performance evaluations by those working most closely with the agency. Obviously, agency performance is multifaceted; the impact of vacancies on performance can vary by context. Still, even in the midst of this complexity, a clear correlation emerges.

*How Does Partisanship Shape Government Performance Evaluation?*

One concern with public sector performance evaluation is that what a Democrat considers good performance might be considered bad performance by Republicans and vice versa. In other words, Democrats and Republicans can observe the same agency activities and outputs but not agree on whether it is good performance. If this is the case, this has serious ramifications not only for this project but performance management efforts more generally. One virtue of our approach to measuring performance is that we know whether federal executives rating agency performance are

Republicans, Independents, or Democrats.[12] This allows us to evaluate whether and how much partisanship shapes evaluations of government performance.

In Table 2 we show the results of a simple bivariate regression where the continuous raw agency rating (slider from 1-Not at all effective to 5-Very effective) is the dependent variable and the partisanship of the rater is the independent variable.[13] The coefficient estimates from the first model show that Democrats rated agency performance about 0.13 lower and Republicans 0.17 higher than independents, on average. The results do not tell us if Democrats and Republicans use the scale similarly, if they rank agencies the same but with an overall shift up or down, or if they see some agencies the same and others differently but they do show that Republican respondents rated Trump-era agencies somewhat higher than Democrats. Importantly, the results in Model 2 show that while Republicans rate agency performance higher than Democrats (and Independents) on average, the effect of partisanship does not differ based upon the ideological leaning of the agency. So, Democrats are no more critical of conservative agencies and Republicans are no more critical of liberal agencies. In effect, what we observe is an intercept shift, where Republicans rate Trump-era agencies somewhat higher and Democrats rate Trump-era agencies somewhat lower.

**Table 2. OLS Models of *Raw* Agency Ratings based Upon Self-reported Partisanship of Rater, 2020**

|  | (1) | (2) |
|---|---|---|
| Democrat rater (0,1) | -0.13 | -0.11 |
|  | [0.05]** | [0.05]** |
| Republican rater (0,1) | 0.17 | 0.16 |

---

[12] We have also evaluated whether agencies that are the subject of regular partisan disagreement are rated differently by Democrats and Republicans and include this analysis in Appendix D. Importantly, this disagreement also does not appear to decrease the *uncertainty* of the estimates.

[13] In total, there were 2,360 ratings by Democrats, 886 ratings by Republicans and 569 ratings by respondents that were independents or refused to identify with a party in the survey.

|  |  |  |
|---|---|---|
|  | [0.05]** | [0.05]** |
| Agency ideology (L,C) |  | 0.14 |
|  |  | [0.04]** |
| Democrat rater*Agency ideology |  | 0.02 |
|  |  | [0.05] |
| Republican rater*Agency ideology |  | 0.05 |
|  |  | [0.05] |
| Constant | 0.03 | 0.01 |
|  | [0.04] | [0.04] |
| N | 3,812 | 3,662 |
| Adj. $R^2$ | 0.02 | 0.04 |
| F-statistic | 30.30** | 31.31** |

Note: ** significant at the 0.05 level; * significant at the 0.10 level in two-tailed tests. Dependent variable is agency rating: "How would you rate the overall performance of the following agencies in carrying out their missions?" (1-5). The online survey response is a slider that allows for continuous values. As such, we estimate models via ordinary least squares (OLS). Base category is ratings by respondents that are independent or do not share a party affiliation.

To further unpack the ways that Democrats and Republicans evaluate performance, we generated new agency performance estimates, one set based on the ratings of Democrat federal executives and one set based only on the ratings of Republican executives. To generate the estimates, we replicate the model above, but limit the ratings to Democratic or Republican raters, respectively.[14] We make sure the estimates are on the same scale by fixing the latent performance for three agencies that span the range of performance.[15] We compare the two sets of ratings to see how

---

[14] We also generate estimates where the hierarchy structuring the estimates is partisanship rather than workplace. In the models, there is little difference in the party-level posterior means of the rater slope parameters 0.06 for Democrats and 0.05 for Republicans (the scale is standard normal). So, a one unit increase in latent performance affects Republican and Democratic perception of performance almost identically. The party-level posteriors for the rater intercepts are 0.06 for Democrats and 0.33 for Republicans, confirming that Republicans rate Trump-era agencies more highly on average.

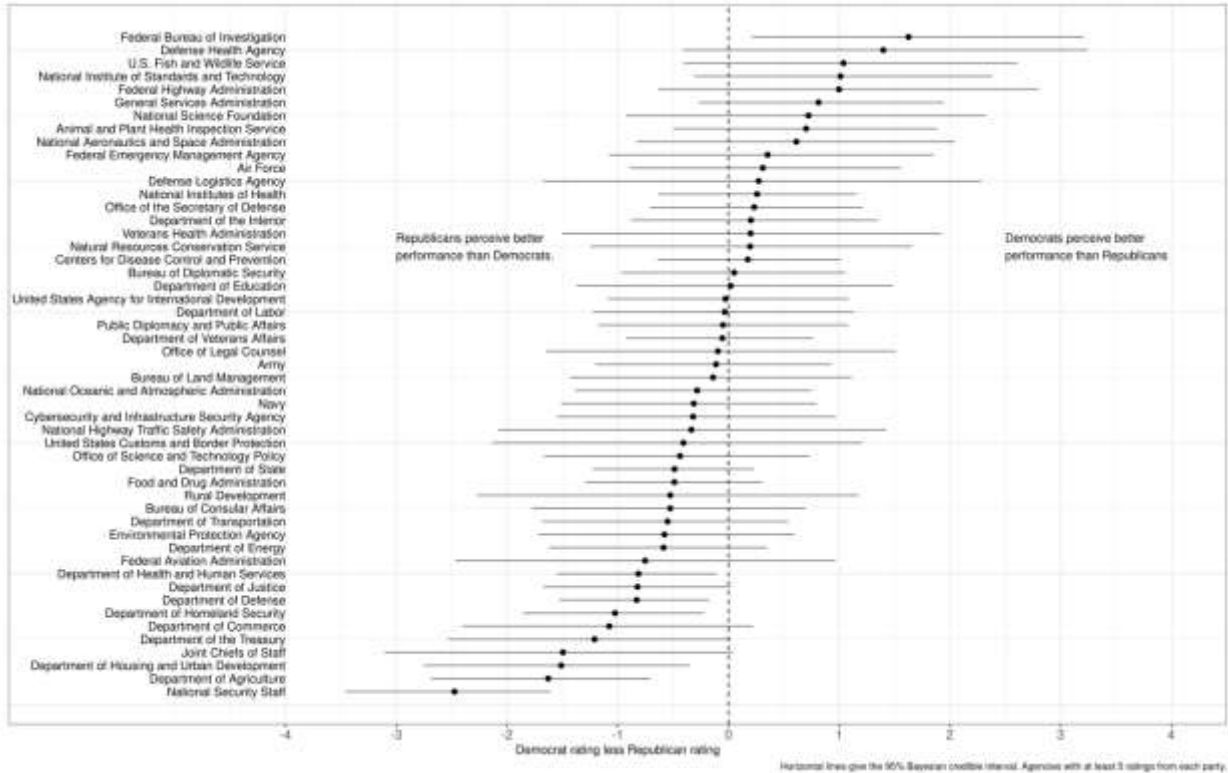[15] For details on how we fix estimates on the same scale see Appendix E.

much the ratings overlap and identify agencies about which Democrats and Republicans agree and

disagree. We generate fewer agency estimates since we have fewer ratings when we separate out

Democrats and Republicans but the results are informative.

**Figure 3. Scatterplot of Agency Performance Estimates with Estimates Generated Separately for Democrat and Republican Raters**
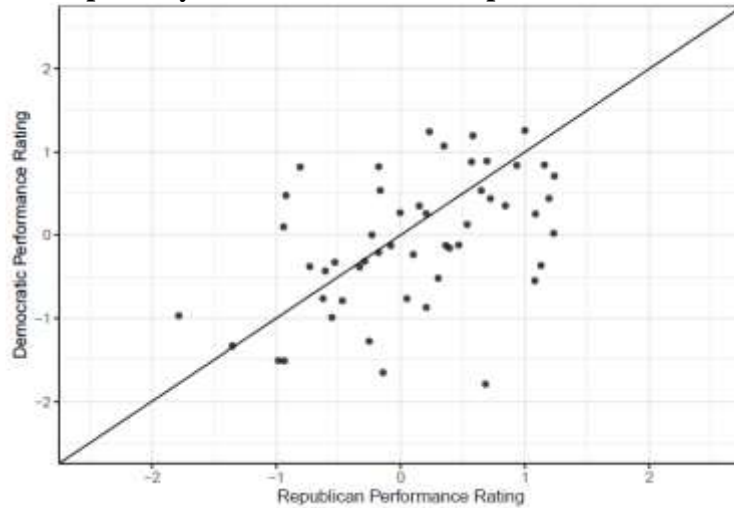


In Figure 3 we include a scatterplot of these estimates with a 45-degree line for reference. In

general, where Republicans rate agencies as high performers, Democrats do the same (corr 0.47).

Where Republicans rate agencies as low performers, Democrats similarly see low performance. This

should give us confidence in the regression estimates above showing a relationship between vacancy

length and performance. Any differences between how Democrats and Republicans evaluate

performance do not lead to materially different their evaluations of underlying performance. This

similarity should also provide confidence to those engaged in ongoing efforts to measure federal

agency performance.

**Figure 4. Difference in Selected Agency Performance Estimates by Party**

Horizontal lines give the 95% Bayesian credible interval. Agencies with at least 3 ratings from each party.

The results thus far suggest that, while Democrats and Republican do sometimes evaluate performance differently (i.e., some dots are well off the 45-degree line), they observe performance the same way, on average. High performers for Democrats are generally high performers for Republicans and vice versa. While this is generally true, there are cases where raters from the two parties tend to disagree. In Figure 4 we graph the difference in agency performance estimates in order to identify the agencies where the Democratic and Republican ratings are notably different. In essence, we are just measuring how far off the 45-degree line an agency is from Figure 3 and including the related statistical uncertainty (i.e., do we have high confidence that the ratings really are different?). In this figure, perfect agreement between Democrats and Republicans means points close to zero. Points further to the right are cases where Democrat federal executives rated an agency more highly than Republican federal executives. On the left are cases where the reverse is true, Republicans rated an agency more highly than Democrats.

A few things stand out in the figure. First, more points are to the left of 0, consistent with our earlier finding that Republicans rated agencies slightly more highly during the Trump Administration than Democrats. Second, most of the points are generally centered around 0, suggesting similar ratings by Democrats and Republicans. Finally, there are only a few cases where we are confident that the ratings differ by party (i.e., the 95% interval does not overlap 0). These include the Departments of Agriculture, Defense, Health and Human Services, Housing and Urban Development, and Homeland Security, as well as the Federal Bureau of Investigation (FBI). Given the political debates of the Trump Administration differing opinions about these agencies are understandable. For example, Democrats argued that the FBI was doing its job when investigating the president and Republicans argued that the president was unfairly targeted. Similarly, efforts to move the Economic Research Service out of Washington likely influenced perceptions of the Performance of the Department of Agriculture. Interestingly, many of the cases where partisans disagree are cabinet departments evaluated as a whole, where identifying a core mission is difficult. In cases where there is no one core mission it is easier for factors like partisanship to define perceptions of performance. Large units often also invite evaluations of the performance based upon opinions of visible leaders (e.g., Ben Carson, William Barr) rather than the agency itself. The partisan differences in the ratings also reveal an underappreciated way that agencies are politicized. In some agencies politics is infused into even into objective evaluations by high level federal executives that evaluate performance as part of their jobs.

To return to our original question, we are concerned about whether vacancies in Senate-confirmed leadership positions harm performance. The results with the basic ratings above suggest that vacancies are harmful for performance, but do partisan differences in ratings influence these results? In Table 3 we replicate the models from Table 1, this time excluding the cases where

Democrats and Republicans disagree significantly on the ratings (Figure 4).[16] The results are broadly

consistent with the estimates reported in Table 1. If anything, the relationship between vacancies

and federal agency performance emerges more clearly. The coefficient estimates are larger and still

estimated precisely, indicating that vacancies in presidentially appointed leadership positions are

correlated with lower agency performance, as measured through the opinions of federal executives

that work outside these agencies. As before, the coefficients on the interaction effects suggest that

the effects of the vacancies taper off the longer a position is vacant. This is important validation that

the main finding from the first section of the paper is robust to partisan differences in those

providing the performance ratings. The breakdown in the appointments process due to slow

nominations and confirmation delays appears to matter for agency performance. Federal executives

that evaluate agency performance on core missions rate agencies with PAS vacancies lower than

agencies with confirmed appointees in place.

**Table 3. OLS Models of the Effect of Vacancies on Agency Performance, Excluding Cases Where Democrats and Republicans Disagree**

| Agency Performance (-2.5, 2.5) | All Agencies | All Agencies | PAS Head Only | PAS Head Only |
|---|---|---|---|---|
| | | | | |
| Months Vacant (0-42) | -0.013 | -0.036 | -0.016 | -0.046 |
| | [0.004]** | [0.010]** | [0.005]** | [0.013]** |
| *Position Controls* | | | | |
| Direct PAS Leader (0,1) | -0.216 | -0.226 | | |
| | [0.138] | [0.132] | | |
| Leadership Transitions (0-4) | 0.105 | 0.123 | 0.121 | 0.145 |
| | [0.033]** | [0.033]** | [0.041]** | [0.039]** |
| *Agency Controls* | | | | |
| EOP (0,1) | -0.731 | -0.671 | -0.717 | -0.647 |

---

[16] Separate estimates for Democrats and Republicans make it possible to estimate models on

Democratic ratings and Republican ratings separately. We do this and discuss the results in

Appendix E.

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| | [0.099]** | [0.088]** | [0.109]** | [0.093]** |
| Cabinet Department (0,1) | 0.401 | 0.422 | 0.424 | 0.442 |
| | [0.075]** | [0.064]** | [0.088]** | [0.076]** |
| Whole Department (0,1) | -0.542 | -0.671 | -0.614 | -0.774 |
| | [0.140]** | [0.151]** | [0.170]** | [0.197]** |
| Independent Commission (0,1) | -0.199 | -0.206 | -0.212 | -0.224 |
| | [0.078]** | [0.070]** | [0.085]** | [0.076]** |
| Priority Department (0,1) | -0.369 | -0.372 | -0.357 | -0.353 |
| | [0.146]** | [0.130]** | [0.168]* | [0.147]** |
| Workforce Skill -- Obama Administration | 0.555 | 0.541 | 0.583 | 0.568 |
| | [0.065]** | [0.067]** | [0.069]** | [0.070]** |
| Agency Ideology (L-C) | 0.037 | 0.025 | -0.019 | -0.04 |
| | [0.081] | [0.083] | [0.098] | [0.100] |
| Months Vacant$^2$ | | 0.0005 | | 0.0007 |
| | | [0.0002]** | | [0.0002]** |
| Constant | 0.368 | 0.500 | 0.151 | 0.304 |
| | [0.158]** | [0.141]** | [0.054]** | [0.059]** |
| $R^2$ | 0.49 | 0.50 | 0.48 | 0.50 |
| N | 116 | 116 | 93 | 93 |

Note: * $p<0.1$; ** $p<0.05$. Dependent variable is agency performance estimate based upon aggregated responses to question: "How would you rate the overall performance of the following agencies in carrying out their missions?" (1-5). Models estimated with Ordinary Least Squares and standard errors are clustered by department. Models 3, 4 estimated using only agencies whose immediate head is a Senate-confirmed political appointee (PAS).

In addition, despite concerns that partisanship endangers the performance measurement enterprise more generally, the evidence here suggests that is not the case. While partisanship comes into play, particularly in larger agencies where mission is unclear and leadership is visible and salient, true latent performance seems to be observed similarly by Democrats and Republicans. While Republicans might have had an easier time seeing the good in Trump Administration agency performance, both Democrats and Republicans appear to see high or low agency performance similarly. In most cases, when Democrats rated agencies more highly, so did Republicans and when Republicans rated an agency more highly, so did Democrats. This does not diminish the importance of the exceptions but suggests that our use of performance measures be tempered by the clarity of what is being measured and its connection to visible political leaders.

**Discussion and Conclusion**

This paper has sought to test the claims made at the start of the Biden Administration that vacancies harm federal agency performance. We did so with original new perceptual measures of U.S. federal agency performance that are comparable across contexts. The paper includes a number of key findings. First, consistent with several recent works, the paper shows that vacancies are associated with lower agency performance (O'Connell 2009, 2020; Piper and Lewis 2023; Rutherford et al. 2018). Second, the effect of a vacancy diminishes over time, suggesting that agencies may adjust to operating without confirmed appointed leadership or, at minimum, stem the decline in performance. Finally, the results are robust to partisan differences among raters.

The focus on the Trump and Biden Administration raises the natural questions of generalizability. The mechanisms at work – short time horizons, less stakeholder investment, and lower morale—are at work in vacancies in other contexts. More work, however, is necessary to unpack when these mechanisms are mitigated by the quality of acting officials or by contexts such as those that make acting officials' tenures durable. Indeed, there is suggestive evidence here that agencies regularly experiencing long vacancies adjust and begin to operate like agencies run by career professionals. If high quality managers serve and expect to be in charge for a while, they can take on long-term problems and persuade others to help them do it. There are other contextual factors here that may amplify or mitigate the effect of vacancies. For example, purposefully chosen political actings with White House backing may drive improved performance (Kinane 2021). Some agencies have multiple and overlapping vacancies and this could alter the impact of any one vacancy. So, while this study provides evidence from one administration, it provides evidence of more general patterns and avenues for future research.

Several other important implications result from these analyses. First, the results highlight the potential consequences of the current dysfunction in the appointments process in the United States. As partisan polarization leads to slower nominations and longer and more uncertain

confirmations (Bond et al. 2009), the results here suggest that the effect will likely be lower agency performance for the foreseeable future. The Senate has changed its rules multiple times in the last two decades to reduce the vacancy problem. They have privileged some nominations by allowing them to skip committee referral, eliminated confirmation requirements for more than 160 positions, shortened debate, and ended supermajority votes for nominations. However, Congress continues to create new Senate confirmed positions and the pace of nominations still lags, caught up in partisan fights leading to objections to unanimous consent and more roll call votes (O'Connell 2021). In 2023, it is taking the Senate twice as long to confirm the nominees as it did in the 1980s (Hitlin et al. 2023).

Second, the performance measurement method employed here is generalizable to other contexts, notably cases where persons outside an organization can be surveyed about their impression of the performance of the organization (Richardson et al. 2018). This is particularly useful in contexts where other objective measures of performance (e.g., profit) are difficult to obtain or assess (Wilson 1989). Data on the persons doing the rating can be used to directly evaluate the quality or potential bias among those evaluating performance, reducing concerns about bias in performance measurement (Lavertu and Moynihan 2013).

Finally, these results suggest that practitioners using performance measures can do so reasonably in many contexts. Indeed, since the vast majority of government work is apolitical or popular with the public, Democrats and Republicans observe performance similarly (Gramlich 2017; Light 2008). Nonetheless, the results here suggest that users of performance measures need to be attuned to potential partisan bias in ratings, particularly as agencies get larger and it becomes harder to identify what core task is being evaluated (Lavertu and Moynihan 2013). In our results, raters evaluating whole departments rated them lower than other agencies and they also disagreed the most

about these ratings. Together, these findings suggest that performance evaluation is easiest and least

subject to bias at lower levels of aggregation and when the performance to be evaluated is clear.

## References

Bellodi, Luca. 2022. "A Dynamic Measure of Bureaucratic Reputation: New Data for New Theory." *American Journal of Political Science*, forthcoming (https://doi.org/10.1111/ajps.12695).

Bernstein, Jonathan. 2022. "A Senseless Logjam is Holding Up Nominations." *Washington Post,* February 9, 2022.

Bond, Jon R., Richard Fleisher, and Glen S. Krutz. 2009. "Malign Neglect: Evidence that Delay Has Become the Primary Method of Defeating Presidential Appointments." *Congress & the Presidency* 36:226-43.

Boylan, Richard T. 2004. "Salaries, Turnover, and Performance in the Federal Criminal Justice System." *The Journal of Law and Economics* 47(1): 75–92.

Boyne, George, and Jay Dahya. 2002. "Executive Succession and the Performance of Public Organizations." *Public Administration* 80(1):179-200.

Boyne, George A., Oliver James, Peter John, Nicolai Petrovsky. 2011. "Top Management Turnover and Organizational Performance: A Test of a Contingency Model." *Public Administration Review* 71(4):572-81.

Dull, Matthew, and Patrick S. Roberts. 2009. "Continuity, Competence, and the Succession of Senate-Confirmed Agency Appointees, 1989-2009." *Presidential Studies Quarterly* 39(3):432–53.

Dull, Matthew, Patrick S. Roberts, Michael S. Keeney, and Sang Ok Choi. 2012. "Appointee Confirmation and Tenure: The Succession of U.S. Federal Agency Appointees, 1989-2009." *Public Administration Review* 72(6):902-12.

Eilperin, Juliet, Josh Dawsey, and Seung Min Kim. 2019. " 'It's way too many': As vacancies pile up in Trump administration, senators grow concerned," *Washington Post*, February 4, 2019.

Farhrenthold, David. 2014. "A Sinkhole of Bureaucracy." *Washington Post*, March 22, 2014.

Fernandez, Sergio, William G. Resh, Tima Moldogaziev, and Zachary W. Oberfield. 2015. "Assessing the Past and Promise of the Federal Employee Viewpoint Survey for Public Management Research: A Research Synthesis." *Public Administration Review* 75(3): 382–94.

Gramer, Robbie. 2022. "3 Years Later, the U.S. Could Finally Send and Ambassador to Ukraine," *Foreign Policy*, April 25, 2022.

Gramlich, John. 2017. "Few Americans support cuts to most government programs, including Medicaid," Pew Research, May 26, 2017 (https://www.pewresearch.org/fact-

tank/2017/05/26/few-americans-support-cuts-to-most-government-programs-including-medicaid/).

Heclo, Hugh. 1977. *A Government of Strangers: Executive Politics in Washington*. Brookings Institution.

Hitlin, Paul, Mary-Courtney Murphy, Carlos Galina. 2023. "Outcome of Midterms Unlikely to Improve Senate Confirmation Process for Executive Branch Nominees." Partnership for Public Service, January 17, 2023 (https://ourpublicservice.org/blog/outcome-of-midterms-unlikely-to-improve-the-senate-confirmation-process/).

Hur, Yongbeom. 2013. "Turnover, Voluntary Turnover, and Organizational Performance: Evidence from Municipal Police Departments." *Public Administration Quarterly* 37(1):3-35.

Hussein, Fatima, Ricardo Alonso-Zaldivar, Hope Yen, and Colleen Long. 2022. "Biden agency vacancies to drag on White House priorities," *ABC News*, February 21, 2022.

Kinane, Christina M. 2021. "Control without Confirmation: The Politics of Vacancies in Presidential Appointments." *American Political Science Review* 115(2):599-614.

Krause, George A., and James W. Douglas. 2006. "Does Agency Competition Improve the Quality of Policy Analysis? Evidence from OMB and CBO Fiscal Projections." *Journal of Policy Analysis and Management* 25(1): 53–74.

Kumar, Martha Joynt. 2021. "Joseph Biden's Effective Presidential Transition: 'Started Early, Went Big' ". *Presidential Studies Quarterly* 51(3): 582-608.

Lavertu, Stephane, and Donald P. Moynihan. "Agency Political Ideology and Reform Implementation: Performance Management in the Bush Administration." *Journal of Public Administration Research & Theory* 23(3):521-49.

Lewis, David E. 2007. "Testing Pendleton's Premise: Do Political Appointees Make Worse Bureaucrats?" *Journal of Politics* 69(4):1073-88.

Light, Paul. 2008. *A Government Ill Executed: The Decline of the Federal Service and How to Reverse It*. Cambridge, MA: Harvard University Press.

Meier, Kenneth J., and Alisa Hicklin. 2007. "Employee Turnover and Organizational Performance: Testing a Hypothesis from Classical Public Administration.*" Journal of Public Administration Research & Theory* 18(4):573-90.

Meier, Kenneth J., and Laurence J. O'Toole Jr. 2002. "Public Management and Organizational Performance: The Effect of Managerial Quality." *Journal of Policy Analysis & Management* 21(4): 629–43.

Meier, Kenneth J., Søren C. Winter, Laurence J. O'Toole, Jr., Nathan Favero, Simon Calmar Andersen. 2015. "The Validity of Subjective Performance Measures: School Principals in Texas and Denmark." *Public Administration* 93(4): 1084–1101.

Mendelson, Nina A. 2014. "The Uncertain Effects of Senate Confirmation Delays in the Agencies." *Duke Law Journal* 64: 1571.

Mendelson, Nina A. 2020. "The Permissibility of Acting Officials: May the President Work Around Senate Confirmation?" *Administrative Law Review* 72(4):533-606.

Moore, Emily H. 2018. "Polarization, Excepted Appointments, and the Administrative Presidency." *Presidential Studies Quarterly* 48(1):72-92.

Moynihan, Donald P., and Sanjay K. Pandey. 2005. "Testing How Management Matters in an Era of Government by Performance Management." *Journal of Public Administration Research & Theory* 15(3):421-39.

Nyhan, Ronald C., and Herbert A. Marlowe Jr. 1995. "Performance Measurement in the Public Sector: Challenges and Opportunities." *Public Productivity & Management Review*: 333–48.

O'Connell, Anne Joseph. 2009. "Vacant Offices: Delays in Staffing Top Agency Positions." *S. Cal. L. Rev.* 82: 913.

———. 2020. "Actings." *Columbia Law Review* 120(3): 613–728.

———. 2021. "The Biden administration still has a lot of vacant positions. It's a growing problem." *Washington Post*, October 28, 2021.

Park, Jungyeon. n.d. "How Individual and Organizational Sources of Managerial Capacity Shape Agency Performance: Evidence from the Size of Improper Payment in U.S. Federal Programs." *Working Paper*.

Partnership for Public Service. 2020. Best Places to Work in the Federal Government Rankings (https://bestplacestowork.org/rankings/?view=category&size=sub&category=leadership&)

Partnership for Public Service. 2002. Political Appointee Tracker, May 6, 2022 (https://ourpublicservice.org/performance-measures/political-appointee-tracker/).

Pfiffner, James P. 1996. *The Strategic Presidency: Hitting the Ground Running*. Lawrence: University Press of Kansas.

Piper, Christopher. 2022. "Presidential Strategy amidst the 'Broken' Appointments Process." *Presidential Studies Quarterly* 52(4): 843-874.

Piper, Christopher, and David E. Lewis. 2023. "Do Vacancies Hurt Federal Agency Performance?" *Journal of Public Administration Research and Theory* 33(2):313-28.

Resh, William G., Gary E. Hollibaugh, Patrick S. Roberts, and Matthew M. Dull. 2021. "Appointee Vacancies in US Executive Branch Agencies." *Journal of Public Policy* 41(4): 653–76.

Richardson, Mark D., Joshua D. Clinton, and David E. Lewis. 2018. "Elite Perceptions of Agency Ideology and Workforce Skill." *The Journal of Politics* 80(1): 303–8.

Rutherford, Amanda. 2016. "The Effect of Top-Management Team Heterogeneity on Performance in Institutions of Higher Education." *Public Performance & Management Review* 40(1): 119–44.

Rutherford, Amanda, and Tama Hameduddin. 2022. "Vacancies Among Appointees in U.S. Federal Agencies: Implications for Employee Attitudes and Intentions." *Governance*, forthcoming.

Rutherford, Amanda, Jeryl Mumpower, Ricardo A. Bello-Gomez, and Marlisa Griffin. 2018. "Understanding Vacancy Time: A Theoretical Framework Informed by Cross-sector Comparison." *Perspectives on Public Management and Governance* 2(1):3-20.

Risen, Clay. 2021. "Why Are So Many Government Positions Still Vacant?" *New York Times*, August 31, 2021.

Stier, Max. 2021. "It's time to fix the broken Senate confirmation process," *The Hill*, November 15, 2021.

Thompson, James R., and Michael D. Siciliano. 2021. "The 'Levels' Problem in Assessing Organizational Climate: Evidence from the Federal Employee Viewpoint Survey." *Public Personnel Management* 50(1):133-56.

Weisner, Molly. 2022. "Federal Retirement Claims Surge, Extending Processing Times." *Federal Times*, February 24, 2022.

Wilson, James Q. 1989. *Bureaucracy*. Basic Books.

Wood, Abby K., and David E. Lewis. 2017. "Agency Performance Challenges and Agency Politicization." *Journal of Public Administration Research and Theory* 27(4): 581–95.

Wrightman, G. Breck, Sergio Fernandez, and Amanda Rutherford. 2022. "Job Vacancy and Organizational Performance: Are Senior Managers or Street-level Bureaucrats Missed Most?" *Public Administration Review* 82(4): 660-70.

*The data underlying this article are available in [Harvard Dataverse Link to come]*

**Appendix A. Details of Estimates of Agency Performance**

To avoid the potential bias associated with self-reports, we use federal executive survey respondents as expert informants to tell us about agencies they work with. At the beginning, each survey instrument asks respondents identify the agencies they work with the most (other than their own). Each respondent was provided three dropdown menus from which to select agencies. The menus included more than 200 agencies organized into 1) the Executive Office of the President, 2) the executive departments and their large sub-components, and 3) independent agencies. So, for example, a respondent might scroll past agencies in the Executive Office of the President and see the Department of Agriculture with 13 options, including "Department of Agriculture (All)" and key bureaus inside the USDA such as the Forest Service or Natural Resources Conservations Service. They could scroll past other departments and their sub-components and get to a list of independent agencies. A screenshot from the survey is included in Figure A1 below.

The survey subsequently includes questions about these agencies as a partial way of getting around problems inherent in self-reports. Respondents were provided 5 agencies to rate. The list of agencies was populated in the following steps:

1. Agencies the respondent selected as one of the three agencies they work with most (up to 3)

2. If the respondent worked in an executive department or the Executive Office of the President, two randomly selected bureaus from the same department or EOP that were not selected by Step 1. If there were fewer than two bureaus remaining after eliminating bureaus from Step 1, then all remaining bureaus were selected. (The Federal Energy Regulatory Commission was treated as bureau of the Department of Energy and the U.S. Agency for International Development was treated as a bureau of the Department of State.)

3. Remaining slots were filled by randomly selected executive departments and independent agencies not selected in Step 1.

The Java code used to populate the list of agencies would not run if the respondent was using Internet Explorer. Therefore, the following process was used to populate the list of agencies for these respondents:

1. Agencies the respondent chose as one of the three agencies they worked with most (up to 3)

2. Office of Management and Budget (if not selected in Step 1)

3. Office of Personnel Management (if not selected in Step 1)

These respondents would be offered fewer than 5 agencies to evaluate if they provided fewer than 3 agencies they work with or selected the Office of management and Budget or Office of Personnel Management as an agency they worked with.

**Figure A1: Question about Which Agencies Respondents Work with Most**

Q4. Please select the three agencies you have worked with the most in order of how often you work with them.
First:
[Drop down menu]
Second:
[Drop down menu]
Third:
[Drop down menu]

Performance: To evaluate whether the independent agency design, in fact, contributes to a more expert workforce both surveys asked respondents the following question (Figure A2):

**Figure A2: Screen Shot of (Online) Agency Performance Question**

**Q38.** How would you rate the overall performance of the following agencies in carrying out their missions?

| | Not at all effective | | | | Very effective | Don't know |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | |
| [Agency 1] | | | [Slider] | | | ○ |
| [Agency 2] | | | [Slider] | | | ○ |
| [Agency 3] | | | [Slider] | | | ○ |
| [Agency 4] | | | [Slider] | | | ○ |
| [Agency 5] | | | [Slider] | | | ○ |

## Estimation of Ratings

We estimate latent performance using a Bayesian multi-rater item response model. Let Y be the I x J ratings matrix with element $y_{ij}$ denoting agency I's evaluation by respondent j. We have ratings for 179 agencies (i.e., I=179), 1,397 raters, and 4,555 ratings. After estimation, we retain only agencies with at least 5 ratings, leaving 142 agencies in the final dataset. We rescaled ratings to be distributed $N(0,1)$ prior to estimating the model to improve model efficiency (see Stan User's Guide Version 2.28, Section 24.12).

We assume an evaluation of agency $i$ by respondent $j$ in workplace $d$ is a function of the agency's latent performance: $y_{ij} \sim N(\alpha_{jd} + \beta_{jd}x_i, \sigma^2)$, where $x_i$ is latent performance. We define workplace as the Executive Office of the President, each executive department, and each independent agency. This response model allows each respondent to have a unique mapping (i.e., location and scale) from their perception of performance to the survey scale. The intercept term, $\alpha_{jd}$, denotes the performance level that respondent $j$ assigns to an agency with performance level $x_i = 0$ on the latent scale. This parameter allows location shifts in the mapping from $x_i$ to $y_{ij}$ across respondents. The parameter $\beta_{jd}$ allows the relationship between the scale of $x_i$ and $y_{ij}$ to vary across respondents. The $\sigma^2$ parameter assumes a common error in respondents' perception of latent performance. An assumption of common error is necessary to prevent divergent transitions

43

caused by an agency-specific error approaching zero, which causes model parameters to get "stuck" as the chain explores the posterior distribution.

We assign $\alpha_{jd}$ and $\beta_{jd}$ a hierarchical prior and assume the parameters for raters who share a workplace are drawn from a common distribution. We have at most 5 ratings per rater and often fewer. The hierarchical prior and the resultant partial pooling across raters improves the precision of the rater parameters and allows us to retain raters who rate only 1 or 2 agencies despite the very limited information we have to estimate their mapping from latent space to the survey scale. Formally, we assume $\alpha_{jd} \sim N(\mu_{d\alpha}, \sigma^2_{d\alpha})$ and $\beta_{jd} \sim N(\mu_{d\beta}, \sigma^2_{d\beta})$ where $d$ indexes workplaces and the $\alpha$ and $\beta$ subscripts denote the corresponding rater parameter. We also estimate models that do not group raters by workplace to examine the implications of this hierarchical model specification. See the relevant section below.

We used informed priors on latent performance to give additional weight to collective perceptions of respondents who reported that an agency is one of the three agencies that they work with most. Formally, we assume $x_i \sim N(\mu_i, \tau^2)$, where $\mu_i$ is the mean evaluation of agency $i$ by respondents who report working with that agency and $\tau^2_i$ is the variance of those ratings. We use all respondents who reported working with an agency to construct the informed prior, including raters who do not meet the two-rating threshold. If an agency has no informed prior, we set $\mu_i = 0$ which is the center of the scale. If an agency has fewer than 5 informed ratings, we set $\tau^2_i = 2.17$, the maximum observed variance of informed ratings for agencies in the model. We also estimate a model without the informed prior, assuming $x_i \sim N(0, 2.25)$ for all agencies. We set the variance to 2.25 so that the prior is near but slightly weaker than the weakest prior used to estimate the informed ratings to facilitate evaluation of the effect of the informed priors on posterior distributions of latent performance. The standard deviations of the posterior distributions for

agencies with few ratings are sensitive to the choice of prior. The posterior distribution of latent

performance was rescaled after estimation to be distributed N(0, 1) to locally identify the scale. The

complete model parameterization with informed priors is:

$$
\begin{aligned}
y_{ij} &\sim N(\alpha_{jd} + \beta_{jd}x_i, \sigma^2) \\
x_i &\sim N(\mu_i, \tau_i^2) \\
\alpha_j &\sim N(\mu_{d\alpha}, \sigma_{d\alpha}^2) \\
\beta_j &\sim \text{half-}N(\mu_{d\beta}, \sigma_{d\beta}^2) \\
\mu_{d\alpha} &\sim N(0,25) \\
\mu_{d\beta} &\sim \text{half-}N(0,25) \\
\sigma^2 &\sim \text{half-}N(0,25) \\
\sigma_{d\alpha}^2 &\sim \text{half-}N(0,25) \\
\sigma_{d\beta}^2 &\sim \text{half-}N(0,25),
\end{aligned}
$$

where $\mu_i$ is the mean of $y_{ij}$ calculated using only informed ratings of agency $i$ and $\tau_i^2$ is the variance

of $y_{ij}$ calculated using only informed ratings of agency $i$. The naive model is identical, except for the

choice of prior on latent performance: $x_i \sim N(0,2.25)$.

The $\beta_j, \mu_{d\beta}, \sigma^2, \sigma_\alpha^2$, and $\sigma_\beta^2$ parameters have a half-normal prior distribution because we

gave them a lower bound of 0. For $\beta_j$ and $\mu_{d\beta}$, this constraint prohibits larger values of latent

performance from mapping to lower ratings on the response scale. This constraint also ensures

rotational identification of the latent scale across chains. The $\sigma^2, \sigma_\alpha^2$, and $\sigma_\beta^2$ parameters are given a

lower bound of 0 because variance parameters are non-negative by definition. The $\alpha_{jd}$ and $\beta_{jd}$

parameters are given non-centered parameterizations. Non-centered parameterizations tend to be

more efficient in hierarchical models. See Stan User's Guide Version 2.30 Section 25.7 for discussion

of hierarchical models and non-centered parameterizations.

We used rstan version 2.21.3 to estimate the model. We ran 5 chains for 4,000 iterations with

the first 1,000 used for warm up. This left 15,000 iterations for inference (3,000 per chain). We used

overdispersed initial values for the latent trait parameters - the first chain was initialized at -3, the

second at -2, the third at 0, the fourth at 2, and the fifth at 3 - to improve the robustness of

inference. Other parameters were initialized at random locations using Stan's default settings. There

were no divergent transitions and model diagnostics indicated the chains converged and the

parameters have sufficient effective samples sizes to make reliable inferences.

## Appendix B. Numerical Estimates of Agency Performance

| Acronym | Agency | Perf. Rating | Standard Deviation | Lower Bound | Upper Bound | # Inf. Ratings | # Ratings |
|---|---|---|---|---|---|---|---|
| ACF | Admin. for Children & Families | 0.30 | 0.30 | -0.28 | 0.90 | 18 | 24 |
| ARPA-E | Adv. Res. Projects Agency-Energy | 0.39 | 0.85 | -1.24 | 2.18 | 2 | 5 |
| AMS | Agricultural Marketing Service | -0.11 | 0.56 | -1.34 | 0.91 | 6 | 10 |
| ARS | Agricultural Research Service | 0.46 | 0.54 | -0.62 | 1.51 | 9 | 13 |
| USAF | Air Force | 0.88 | 0.26 | 0.38 | 1.42 | 37 | 40 |
| APHIS | Animal & Plant Health Ins. Service | 0.58 | 0.30 | 0.01 | 1.20 | 28 | 34 |
| T | Arms Control & Int. Security Affairs | -0.64 | 0.31 | -1.28 | -0.05 | 7 | 18 |
| ARMY | Army | 0.68 | 0.21 | 0.29 | 1.11 | 48 | 49 |
| BCA | Bureau of Consular Affairs | 0.91 | 0.24 | 0.48 | 1.42 | 10 | 39 |
| BDS | Bureau of Diplomatic Security | 0.33 | 0.21 | -0.07 | 0.76 | 7 | 42 |
| BEA | Bureau of Economic Analysis | 1.07 | 0.57 | 0.03 | 2.28 | 3 | 9 |
| BIA | Bureau of Indian Affairs | -1.43 | 0.38 | -2.23 | -0.74 | 12 | 17 |
| BINLEA | Bur. of Int. Narc. & Law Enf. Affairs | 0.19 | 0.28 | -0.34 | 0.76 | 3 | 25 |
| BLS | Bureau of Labor Statistics | 1.20 | 0.39 | 0.47 | 2.02 | 13 | 14 |
| BLM | Bureau of Land Management | -0.91 | 0.28 | -1.52 | -0.41 | 16 | 26 |
| BOEM | Bureau of Ocean Energy Management | 0.40 | 0.45 | -0.43 | 1.35 | 4 | 7 |
| BOR | Bureau of Reclamation | 0.53 | 0.39 | -0.22 | 1.32 | 7 | 10 |
| BFS | Bureau of the Fiscal Service | 0.22 | 0.26 | -0.26 | 0.77 | 13 | 13 |
| CDC | Centers for Disease Control & Prev. | -0.24 | 0.19 | -0.61 | 0.13 | 70 | 94 |
| CMS | Cen. for Medicare & Medicaid Serv. | 0.37 | 0.26 | -0.13 | 0.89 | 24 | 34 |
| CIA | Central Intelligence Agency | 0.88 | 0.25 | 0.42 | 1.39 | 22 | 24 |
| CIS | Citizenship and Immigration Services | -0.12 | 0.43 | -0.95 | 0.75 | 5 | 10 |
| J | Civ. Security, Dem., & Human Rights | -0.62 | 0.31 | -1.26 | -0.05 | 5 | 23 |
| USCG | Coast Guard | 1.16 | 0.29 | 0.63 | 1.78 | 25 | 32 |
| CFPB | Consumer Financial Protection Bur. | -0.65 | 0.43 | -1.59 | 0.12 | 8 | 9 |
| CEA | Council of Economic Advisers | -1.82 | 0.48 | -2.87 | -0.96 | 6 | 9 |
| CEQ | Council on Environmental Quality | -0.43 | 0.53 | -1.54 | 0.54 | 11 | 11 |
| CBP | Customs and Border Protection | -0.33 | 0.31 | -0.95 | 0.30 | 16 | 21 |
| CISA | Cyber. & Infrastructure Sec. Agency | -0.91 | 0.29 | -1.50 | -0.37 | 21 | 27 |
| DARPA | Def. Adv. Research Projects Agency | 0.92 | 0.40 | 0.14 | 1.74 | 9 | 14 |
| DCA | Defense Commissary Agency | 0.20 | 0.85 | -1.48 | 1.87 | 1 | 5 |
| DCAA | Defense Contract Audit Agency | 0.26 | 0.78 | -1.26 | 1.84 | 1 | 7 |
| DCMA | Def. Contract Management Agency | 0.80 | 0.42 | -0.02 | 1.67 | 6 | 10 |
| DFAS | Def. Finance & Accounting Service | 0.04 | 0.45 | -0.85 | 0.91 | 8 | 15 |
| DHA | Defense Health Agency | -0.10 | 0.35 | -0.80 | 0.60 | 12 | 15 |
| DLA | Defense Logistics Agency | 0.84 | 0.29 | 0.29 | 1.43 | 10 | 15 |
| USDA | Department of Agriculture | -0.10 | 0.24 | -0.58 | 0.36 | 61 | 61 |
| COM | Department of Commerce | -0.67 | 0.20 | -1.09 | -0.28 | 58 | 62 |
| DOD | Department of Defense | 0.57 | 0.14 | 0.31 | 0.86 | 173 | 175 |

| DOED | Department of Education | -1.32 | 0.31 | -1.99 | -0.74 | 25 | 26 |
|---|---|---|---|---|---|---|---|
| DOE | Department of Energy | 0.15 | 0.23 | -0.30 | 0.62 | 49 | 52 |
| HHS | Dep. of Health & Human Services | -0.50 | 0.16 | -0.82 | -0.20 | 135 | 137 |
| DHS | Department of Homeland Security | -0.79 | 0.19 | -1.18 | -0.43 | 83 | 87 |
| HUD | Dep. of Housing & Urban Dev. | -1.36 | 0.30 | -1.99 | -0.82 | 36 | 36 |
| DOJ | Department of Justice | -0.32 | 0.16 | -0.63 | -0.01 | 102 | 109 |
| DOL | Department of Labor | -0.24 | 0.29 | -0.81 | 0.34 | 28 | 30 |
| STAT | Department of State | 0.04 | 0.15 | -0.25 | 0.33 | 131 | 136 |
| INT | Department of the Interior | -0.41 | 0.21 | -0.84 | -0.02 | 45 | 47 |
| TREAS | Department of the Treasury | 0.22 | 0.24 | -0.23 | 0.72 | 36 | 37 |
| DOT | Department of Transportation | -0.12 | 0.23 | -0.59 | 0.33 | 46 | 50 |
| DVA | Department of Veterans Affairs | -0.25 | 0.18 | -0.60 | 0.11 | 58 | 60 |
| DEA | Drug Enforcement Administration | -0.34 | 0.49 | -1.31 | 0.59 | 7 | 8 |
| EDA | Economic Development Admin. | -0.67 | 0.47 | -1.65 | 0.20 | 4 | 8 |
| E | Econ. Growth, Energy, & the Env. | -0.28 | 0.23 | -0.75 | 0.18 | 23 | 41 |
| ERS | Economic Research Service | 0.12 | 0.76 | -1.38 | 1.65 | 5 | 10 |
| ETA | Employment & Training Admin. | 0.34 | 0.43 | -0.47 | 1.21 | 11 | 12 |
| EPA | Environmental Protection Agency | -1.43 | 0.27 | -1.98 | -0.94 | 51 | 51 |
| EEOC | Equal Emp. Opportunity Commission | -0.46 | 0.31 | -1.09 | 0.11 | 12 | 14 |
| EOUSA | Executive Office for U.S. Attorneys | 1.13 | 0.39 | 0.40 | 1.92 | 10 | 14 |
| FSA | Farm Service Agency | -0.14 | 0.51 | -1.16 | 0.87 | 16 | 23 |
| FAA | Federal Aviation Administration | 0.80 | 0.33 | 0.18 | 1.46 | 16 | 22 |
| FBI | Federal Bureau of Investigation | 0.26 | 0.27 | -0.27 | 0.82 | 25 | 28 |
| FDIC | Federal Deposit Insurance Corp. | 0.96 | 0.28 | 0.43 | 1.52 | 7 | 10 |
| FEMA | Federal Emergency Mgt Agency | -0.26 | 0.29 | -0.84 | 0.30 | 29 | 34 |
| FHWA | Federal Highway Administration | 0.26 | 0.33 | -0.38 | 0.95 | 9 | 15 |
| HOU | Fed. Housing Admin./Ofc of Housing | -0.12 | 0.33 | -0.76 | 0.53 | 9 | 19 |
| FLRA | Federal Labor Relations Authority | -0.84 | 0.37 | -1.63 | -0.18 | 3 | 6 |
| FMCSA | Fed Motor Carrier Safety Admin. | -0.45 | 0.59 | -1.66 | 0.70 | 3 | 6 |
| FED | Federal Reserve | 0.95 | 0.36 | 0.28 | 1.70 | 11 | 12 |
| FTC | Federal Trade Commission | 0.75 | 0.50 | -0.11 | 1.84 | 8 | 8 |
| FTA | Federal Transit Administration | -0.07 | 0.40 | -0.85 | 0.71 | 8 | 13 |
| FDA | Food and Drug Administration | 0.68 | 0.21 | 0.29 | 1.12 | 50 | 65 |
| FNS | Food and Nutrition Service | -0.40 | 0.60 | -1.56 | 0.78 | 6 | 7 |
| FSIS | Food Safety and Inspection Service | 0.33 | 0.35 | -0.34 | 1.03 | 18 | 22 |
| FS | Forest Service | 0.37 | 0.36 | -0.32 | 1.09 | 16 | 21 |
| GSA | General Services Administration | -1.48 | 0.27 | -2.04 | -0.98 | 56 | 61 |
| GNMA | Govt National Mortgage Association | 0.93 | 0.51 | 0.10 | 2.12 | 2 | 9 |
| HRSA | Health Resources & Services Admin. | 0.27 | 0.32 | -0.33 | 0.91 | 15 | 20 |
| ICE | Immigration & Customs Enforcement | -1.06 | 0.42 | -1.91 | -0.28 | 13 | 15 |
| HIS | Indian Health Service | -1.07 | 0.40 | -1.91 | -0.32 | 7 | 13 |

| | | | | | | |
|---|---|---|---|---|---|---|
| IES | Institute of Education Sciences | 0.47 | 0.64 | -0.72 | 1.79 | 5 | 6 |
| IRS | Internal Revenue Service | -0.06 | 0.30 | -0.63 | 0.55 | 19 | 21 |
| ITA | International Trade Administration | 0.30 | 0.30 | -0.29 | 0.90 | 18 | 24 |
| JCS | Joint Chiefs of Staff | 0.33 | 0.36 | -0.35 | 1.06 | 14 | 17 |
| M | Management | 0.24 | 0.23 | -0.20 | 0.69 | 10 | 43 |
| MSPB | Merit Systems Protection Board | -1.15 | 0.56 | -2.41 | -0.23 | 4 | 6 |
| MCC | Millennium Challenge Corporation | 0.52 | 0.50 | -0.41 | 1.54 | 6 | 6 |
| MSHA | Mine Safety and Health Administration | 0.50 | 0.54 | -0.51 | 1.65 | 4 | 6 |
| NASA | National Aeronautics & Space Admin. | 1.10 | 0.31 | 0.54 | 1.73 | 22 | 28 |
| NASS | National Agricultural Statistics Service | 1.02 | 0.47 | 0.12 | 1.99 | 10 | 17 |
| NARA | National Archives & Records Admin. | -0.39 | 0.75 | -2.01 | 1.01 | 4 | 5 |
| NCA | National Cemetery Administration | 1.40 | 0.54 | 0.45 | 2.56 | 0 | 8 |
| NEH | Nat. Endowment for the Humanities | -0.32 | 0.63 | -1.58 | 0.92 | 5 | 5 |
| NHTSA | Nat. Highway Traffic Safety Admin. | 0.03 | 0.39 | -0.74 | 0.80 | 9 | 13 |
| NIST | National Institute of Standards & Tech | 0.72 | 0.31 | 0.14 | 1.37 | 15 | 27 |
| NIH | National Institutes of Health | 1.37 | 0.24 | 0.92 | 1.87 | 59 | 74 |
| NNSA | National Nuclear Security Admi. | 0.45 | 0.38 | -0.30 | 1.22 | 15 | 17 |
| NOAA | Nat. Oceanic & Atmospheric Admin. | 0.67 | 0.25 | 0.20 | 1.18 | 37 | 48 |
| NPS | National Park Service | 0.68 | 0.24 | 0.25 | 1.19 | 17 | 28 |
| NSF | National Science Foundation | 1.15 | 0.28 | 0.62 | 1.72 | 37 | 43 |
| NSA | National Security Agency | 1.41 | 0.30 | 0.83 | 2.03 | 7 | 7 |
| NSS | National Security Staff | -1.05 | 0.20 | -1.47 | -0.69 | 96 | 100 |
| NTSB | National Transportation Safety Board | -1.49 | 0.57 | -2.68 | -0.42 | 7 | 10 |
| NRCS | Natural Resources Conservation Ser. | 0.91 | 0.39 | 0.18 | 1.69 | 18 | 25 |
| NAVY | Navy | 0.96 | 0.24 | 0.51 | 1.45 | 42 | 45 |
| NRC | Nuclear Regulatory Commission | 0.47 | 0.77 | -1.00 | 2.03 | 5 | 5 |
| OSHA | Occupational Safety & Health Admin. | -0.32 | 0.58 | -1.55 | 0.75 | 8 | 9 |
| OE | Office of Electricity | 0.35 | 0.63 | -0.99 | 1.55 | 4 | 6 |
| OESE | Office of Elementary & Secondary Ed. | -1.00 | 0.59 | -2.19 | 0.18 | 3 | 7 |
| EERE | Ofc of Energy Eff. & Renewable Ener. | 1.00 | 0.45 | 0.17 | 1.93 | 12 | 14 |
| OEM | Office of Environmental Management | -1.90 | 0.51 | -2.98 | -0.99 | 8 | 8 |
| OFSA | Office of Federal Student Aid | -0.59 | 0.58 | -1.73 | 0.55 | 4 | 8 |
| OGE | Office of Government Ethics | 0.18 | 0.55 | -0.88 | 1.27 | 6 | 7 |
| OLC | Office of Legal Counsel | 1.17 | 0.38 | 0.45 | 1.93 | 12 | 15 |
| OMB | Office of Management and Budget | -1.04 | 0.14 | -1.33 | -0.78 | 180 | 415 |
| ONDCP | Office of National Drug Control Policy | -1.18 | 0.64 | -2.50 | 0.03 | 5 | 6 |
| ONE | Office of Nuclear Energy | -1.41 | 0.79 | -3.05 | 0.12 | 5 | 6 |
| OPM | Office of Personnel Management | -1.96 | 0.21 | -2.39 | -1.57 | 57 | 318 |
| SC | Office of Science | 1.13 | 0.34 | 0.49 | 1.84 | 19 | 20 |
| OSTP | Office of Science & Technology Policy | -1.23 | 0.30 | -1.85 | -0.67 | 36 | 40 |
| OSC | Office of Special Counsel | 0.53 | 0.64 | -0.61 | 1.94 | 4 | 7 |

| | | | | | | |
|---|---|---|---|---|---|---|
| OCC | Off. of the Comptroller of Currency | 0.50 | 0.35 | -0.15 | 1.23 | 8 | 8 |
| ODNI | Office of the Dir. of Nat. Intelligence | 0.13 | 0.28 | -0.41 | 0.70 | 11 | 13 |
| OSD | Office of the Secretary of Defense | -0.07 | 0.19 | -0.45 | 0.30 | 60 | 62 |
| USTR | Office of the U.S. Trade Representative | -0.11 | 0.35 | -0.82 | 0.57 | 26 | 26 |
| OWCP | Ofc of Workers' Compensation Prog. | -0.87 | 0.85 | -2.60 | 0.82 | 4 | 6 |
| P | Political Affairs | 0.45 | 0.26 | -0.04 | 0.97 | 13 | 26 |
| R | Public Diplomacy and Public Affairs | 0.09 | 0.22 | -0.34 | 0.54 | 12 | 39 |
| RD | Rural Development | -0.89 | 0.45 | -1.79 | -0.02 | 20 | 21 |
| USSS | Secret Service | 0.67 | 0.54 | -0.31 | 1.84 | 4 | 8 |
| SBA | Small Business Administration | 0.04 | 0.37 | -0.69 | 0.76 | 18 | 18 |
| SSA | Social Security Administration | -0.44 | 0.34 | -1.06 | 0.26 | 14 | 18 |
| TSA | Transportation Security Administration | -0.68 | 0.45 | -1.58 | 0.18 | 8 | 11 |
| CEN | U.S. Census Bureau | 0.29 | 0.26 | -0.19 | 0.84 | 15 | 26 |
| FWS | U.S. Fish and Wildlife Service | 0.43 | 0.22 | 0.02 | 0.87 | 31 | 39 |
| USGS | U.S. Geological Survey | 0.47 | 0.26 | 0.00 | 1.00 | 27 | 29 |
| PTO | U.S. Patent and Trademark Office | 0.96 | 0.63 | -0.19 | 2.29 | 3 | 9 |
| USAID | U.S. Agency for Int. Development | 0.00 | 0.17 | -0.34 | 0.34 | 56 | 74 |
| USPS | United States Postal Service | -0.09 | 0.71 | -1.52 | 1.30 | 4 | 5 |
| VBA | Veterans' Benefits Administration | -1.07 | 0.39 | -1.87 | -0.34 | 13 | 18 |
| VHA | Veterans' Health Administration | 0.04 | 0.36 | -0.65 | 0.75 | 13 | 17 |
| WHD | Wage and Hour Division | -1.37 | 0.70 | -2.82 | -0.06 | 5 | 6 |

## Appendix C. Comparison of Hierarchical and Non-hierarchical Models of Agency Performance, 2020

We have also estimated models assuming no hierarchical structure and compared these estimates to the hierarchical estimates. The estimates are quite similar.

*Non-hierarchical model: Estimation details*

We assume an evaluation of agency $i$ by respondent $j$ is a function of the agency's latent performance: $y_{ij} \sim N(\alpha_j^* + \beta_j x_i, \sigma^2)$, where $x_i$ is latent performance. This response model allows each respondent to have a unique mapping (i.e., location and scale) from their perception of performance to the survey scale. The intercept term, $\alpha_j^*$, denotes the performance level that respondent $j$ assigns to an agency with performance level $x_i = 0$ on the latent scale. This parameter allows location shifts in the mapping from $x_i$ to $y_{ij}$ across respondents. The parameter $\beta_j$ allows the relationship between the scale of $x_i$ and $y_{ij}$ to vary across respondents. The $\sigma^2$ parameter assumes a common error in respondents perception of latent performance. An assumption of common error is necessary to prevent divergent transitions caused by an agency-specific error approaching zero, which causes model parameters to get "stuck" as the chain explores the posterior distribution.

We used informed priors on latent performance to give additional weight to collective perceptions of respondents who reported that an agency is one of the three agencies that they work with most. Formally, we assume $x_i \sim N(\mu_i, \tau^2)$, where $\mu_i$ is the mean evaluation of agency $i$ by respondents who report working with that agency and $\tau_i^2$ is the variance of those ratings. We use all respondents who reported working with an agency to construct the informed prior, including raters who do not meet the two-rating threshold. If an agency has no informed prior, we set $\mu_i = 0$ which is the center of the scale. If an agency has fewer than 5 informed ratings, we set $\tau_i^2 = 2.17$, the maximum observe variance of informed ratings for agencies in the model. We also estimate a model without the informed prior, assuming $x_i \sim N(0, 2.25)$ for all agencies. We set the variance to 2.25

so that the prior is near but slightly weaker than the weakest prior used to estimate the informed

ratings to facilitate evaluation of the effect of the informed priors on posterior distributions of latent

performance. The standard deviations of the posterior distributions for agencies with few ratings are

sensitive to the choice of prior. The posterior distribution of latent performance was rescaled after

estimation to be distributed $N(0, 1)$ to locally identify the scale. The complete model

parameterization with informed priors is:

$$
\begin{aligned}
y_{ij} &\sim N((\alpha_j + \theta_\alpha) + \beta_j x_i, \sigma^2) \\
x_i &\sim N(\mu_i, \tau_i^2) \\
\alpha_j &\sim N(0, \sigma_\alpha^2) \\
\beta_j &\sim \text{half-}N(0, \sigma_\beta^2) \\
\theta_\alpha &\sim N(0,25) \\
\sigma^2 &\sim \text{half-}N(0,25) \\
\sigma_\alpha^2 &\sim \text{half-}N(0,25) \\
\sigma_\beta^2 &\sim \text{half-}N(0,25),
\end{aligned}
$$

where $\mu_i$ is the mean of $y_{ij}$ calculated using only informed ratings of agency $i$ and $\tau_i^2$ is the variance

of $y_{ij}$ calculated using only informed ratings of agency $i$. The naive model is identical, except for the

choice of prior on latent performance: $x_i \sim N(0,2.25)$.

The $\beta_j, \sigma^2, \sigma_\alpha^2$, and $\sigma_\beta^2$ parameters have a half-normal prior distribution because we gave

them a lower bound of 0. For $\beta_j$, this constraint prohibits larger values of latent performance from

mapping to lower ratings on the response scale. This constraint also ensures rotational identification

of the latent scale across chains. The $\sigma^2, \sigma_\alpha^2$, and $\sigma_\beta^2$ parameters are given a lower bound of 0

because variance parameters are non-negative by definition. The $\alpha_j$ parameter is given a non-

centered parameterization, with grand mean $\theta_\alpha$. Non-centered parameterizations tend to be more

efficient in hierarchical models. In the initial exposition of the model above, we used $\alpha_j^* = \alpha_j + \theta_\alpha$

to simplify presentation. See Stan User's Guide Version 2.30 Section 1.11 for discussion of both the

constraint on $\beta_j$ and non-centered parameterization of $\alpha_j$ for item response models.

We used rstan version 2.21.3 to estimate the model. We ran 4 chains for 4,000 iterations with the first 1,000 used for warm up. This left 12,000 iterations for inference (3,000 per chain). We used overdispersed initial values for the latent trait parameters - the first chain was initialized at -3, the second at -2, the third at 2, and the fourth at 3 - to improve the robustness of inference. Other parameters were initialized at random locations using Stan's default settings. There were no divergent transitions and model diagnostics indicated the chains converged and the parameters have sufficient effective samples sizes to make reliable inferences.

*Comparing Hierarchical and Non-hierarchical Models*

In Figure C1 we compare the posterior means and standard deviations for the hierarchical and non-hierarchical models. Overall, there are not large differences in posterior means. The posterior means are highly correlated at 0.98. The performance ratings on the lower end of the scale, namely below 0, tend to be closer to zero in the hierarchical model than the non-hierarchical model as evidenced by most of the posterior means above the 45-degree line in that portion of the scale. In other words, the lower performing agencies get higher scores in the hierarchical models. Similarly, the standard deviations are similar across models with correlations of 0.94. The hierarchical model yields smaller standard deviations for 51% of agencies in the informed models.

**Figure C1. Comparison of Posterior Means and Standard Deviations for Hierarchical and Non-hierarchical Models**
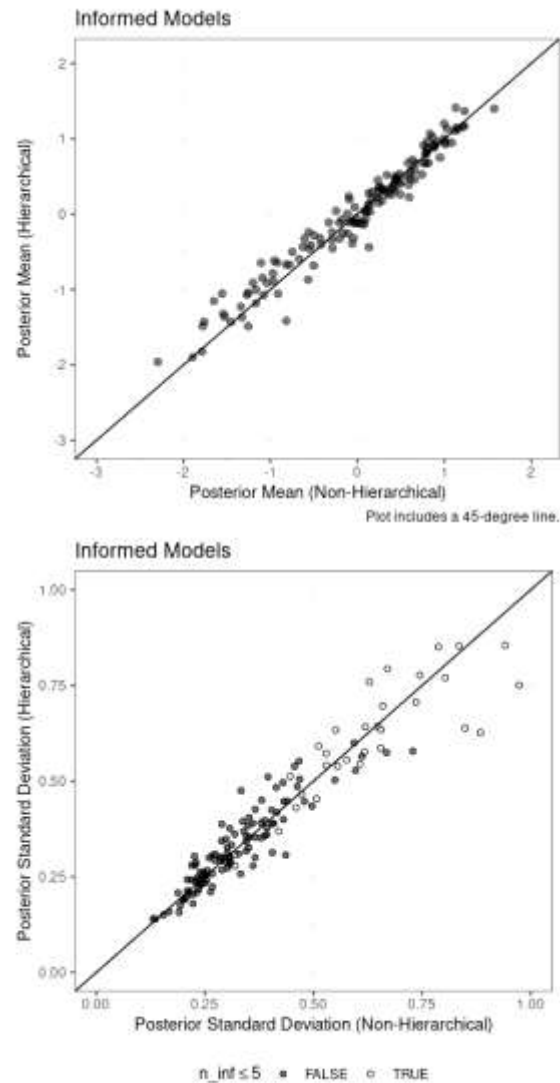


*Comparing Rater Parameters*

Figure C2 compares the posterior means and standard deviations for the rater parameters between hierarchical and non-hierarchical models. The top row shows that the slope parameters (i.e., $\beta$'s) tend to be larger in the hierarchical models. The posterior means of the slope parameters are larger in the hierarchical model for 97% of raters and the median absolute difference is 0.23. Therefore, there is a stronger relationship between the latent performance and the survey scale in the hierarchical model. The standard deviations of the posterior distributions of slope parameters

also tend to be larger in hierarchical model. The standard deviations are larger for 80% of raters and the median absolute difference is 0.058.

The top row of Figure C3 shows that the extreme intercept parameters (i.e., $\alpha$'s) tend to be closer to zero in the hierarchical models as evidenced by ratings below zero tending to be below the 45-degree line and ratings above zero tending to be above the line. Recall the raw ratings are rescaled to be distributed N(0,1) before the Bayesian models are estimated, meaning 0 is the mean of the observed ratings. Therefore, raters with extreme intercepts would assign an agency with latent performance of zero a value closer to zero in the hierarchical model than the non-hierarchical model. In other words, these extreme raters now evaluate performance more similarly to other raters, as we would expect if the hierarchical model makes raters' parameters more similar within groups. Similar to the slope parameters, the standard deviations are larger for 51% of the raters in the hierarchical models than the nonhierarchical informed models.

Overall, the hierarchical model has two effects on rater parameters. First, there is a stronger relationship between latent performance and assigned performance in the hierarchical models (due to the larger slope parameters). Second, raters with extreme thresholds use the scale more similarly to other raters (due to the more similar intercept parameters).

**Figure C2. Comparison of Posterior Means and Standard Deviations for Rater Slope Parameters for Hierarchical and Non-hierarchical Models**
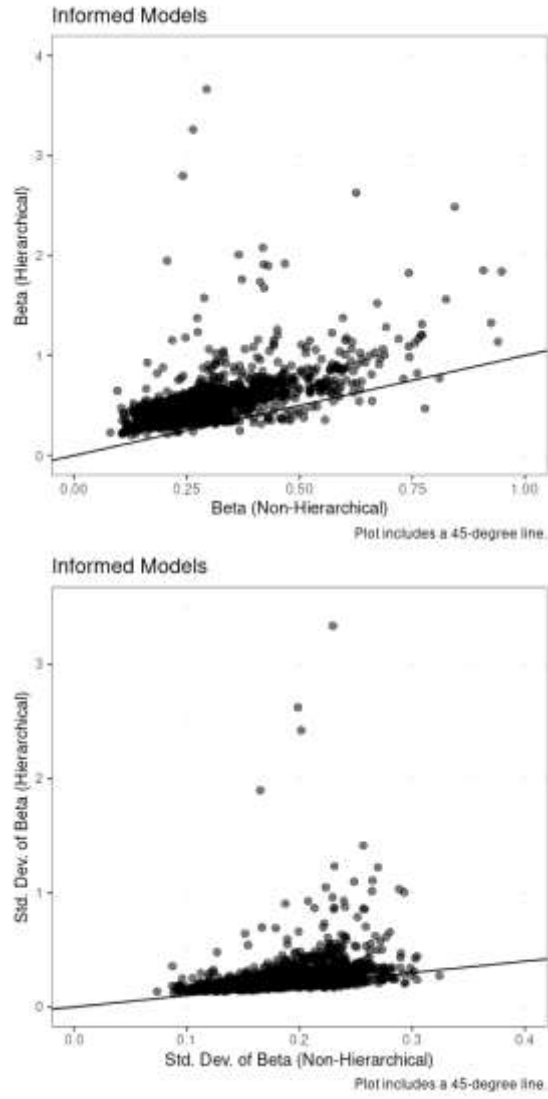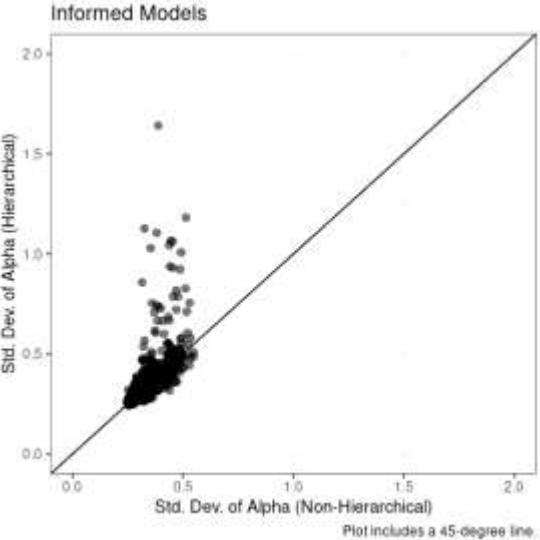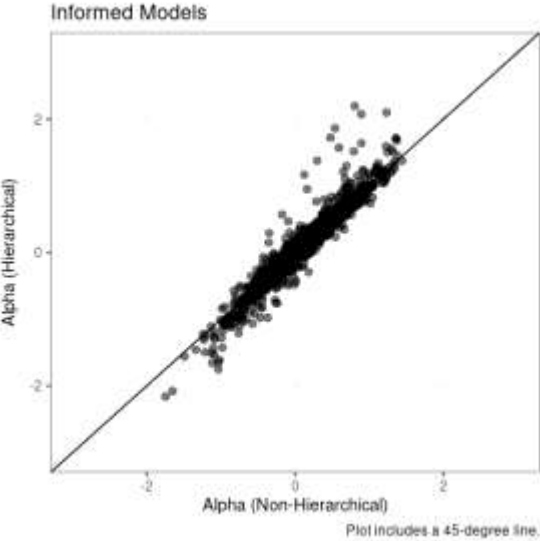
**Figure C3. Comparison of Posterior Means and Standard Deviations for Rater Intercept Parameters for Hierarchical and Non-hierarchical Models**

Informed Models

Plot includes a 45-degree line.

Plot includes a 45-degree line.

## Appendix D. Partisan Disagreement and Agency Performance Evaluations

As a way of evaluating the impact of party on performance evaluations we compare our agency performance ratings to a measure of partisan disagreement to see whether agencies that are subject to regular partisan disagreement are rated differently by Democrats and Republicans. In the 2020 survey we asked respondents "How often do Republicans and Democrats in Washington disagree over what [your agency] should do?" [Never, Rarely, Sometimes, Often, Always, Don't know] We regress the performance ratings on the normalized agency average responses to this question about partisan disagreement and graph these estimates in Figure 2.[17] Interestingly, the more an agency is subject to partisan disagreement, the lower its rating overall. When Democrats and Republicans disagree, they both seem to think the agency is performing poorly. So, rather than Democrats rating agencies high and Republicans rating agencies low or vice versa, it looks like the raters agree on performance.

Importantly, this disagreement also does not appear to decrease the *uncertainty* of the estimates. If Democrats and Republicans saw performance differently when they disagreed, this should increase the uncertainty of the estimates. For example, if Republicans and Democrats always gave 1s and 5s, respectively, when they disagreed and always gave similar scores when they agreed, we should see less precision in the estimates when they disagree. The bottom panel in Figure D1 regresses the standard deviation of the performance estimates on partisan disagreement and shows no relationship. This is evidence that partisan disagreement does not lead to divergent evaluations, at least for most agencies.

---

[17] We rescaled the agency average responses to the partisan disagreement question to be distributed $N(0,1)$.

**Figure D1. Estimates of Agency Performance Ratings based Upon Partisan Disagreement About Agency Work, 2020**
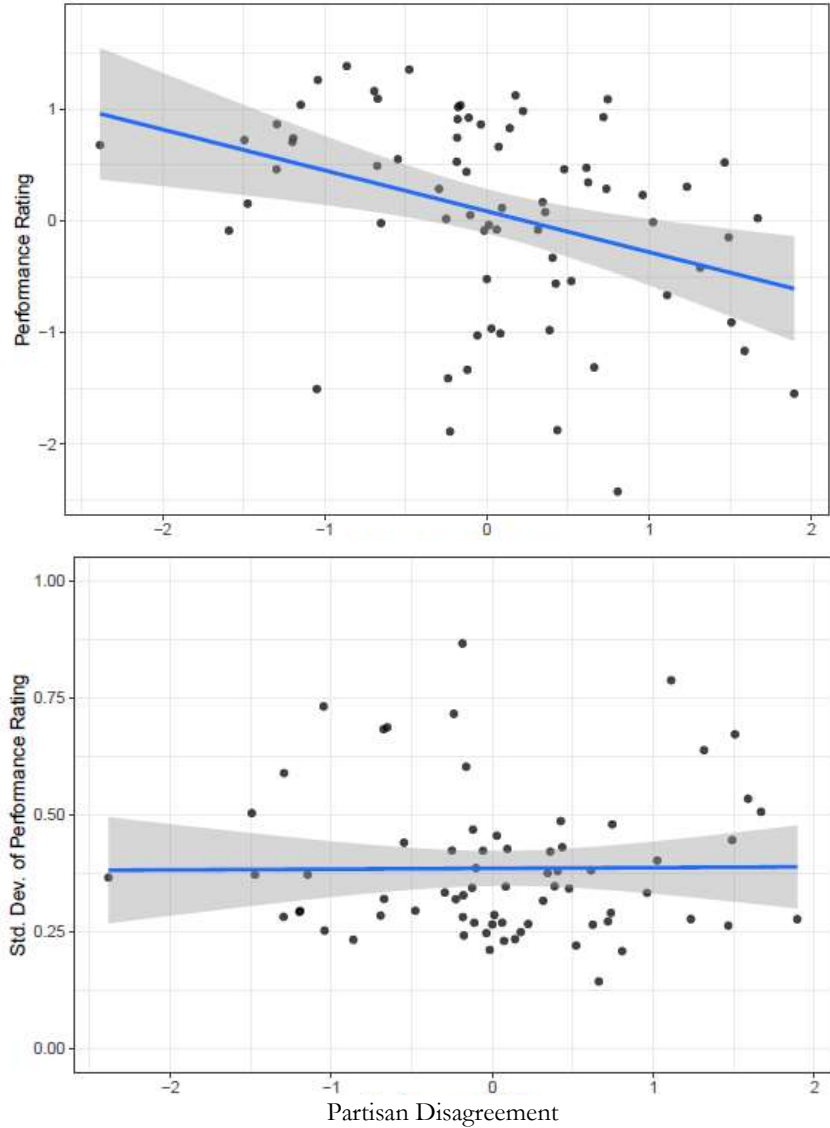


Note: The top panel includes a scatter plot and fitted line of agency performance ratings (y-axis) and agency average responses to the following question about partisan disagreement: "How often do Republicans and Democrats in Washington disagree over what [your agency] should do?" [Never, Rarely, Sometimes, Often, Always, Don't know] (x-axis). The bottom panel includes a scatter plot and fitted line of the standard deviation of agency performance ratings (y-axis) and agency average responses to the partisan disagreement question (x-axis).

**Appendix E. Partisan Performance Ratings**

Separate estimates for Democrats and Republicans make it possible to estimate models on Democratic ratings and Republican ratings separately. This is difficult for two reasons. First, since there are fewer raters, we get fewer agency performance estimates by party and the agencies that remain tend to be larger agencies (e.g., executive departments). Second, given what we know about some of the partisan ratings being determined by partisanship rather than performance, estimating such models would be acknowledging that we are not regressing performance on vacancies. Rather, we are regressing partisan perceptions of performance on vacancies. One solution is to estimate separate models on Democrat and Republican ratings but exclude the known cases of partisan disagreement. This leaves us with 37 cases.

To generate estimates separate estimates for Republican and Democratic raters we need a way that fixes the scale. We used three criteria to identify the agencies for which we fixed their latent performance across parties to recover separate performance ratings by Republicans and Democrats on the same scale for other agencies. First, we wanted agencies whose raw means were very similar across parties suggesting that Republicans and Democrats do indeed perceive performance similarly. Second, we wanted agencies that span the range of performance with one agency at the low end of the scale and one at the high end to locate the ends of the scale. Third, we wanted agencies that many respondents from each party rated to anchor interpersonal comparisons across parties for as many raters as possible. Using these criteria, we selected the Office of Management and Budget, the Office of Personnel Management, and the Coast Guard. Each agency has similar mean performance among Democrats and Republicans. Specifically, the mean rating of OMB on the 1-5 scale was 3.23 among Republicans and 3.12 among Democrats, mean rating of OPM was 2.84 among Republicans and 2.57 among Democrats, and the mean rating of the Coast Guard was 4.35 among Republicans and 4.07 among Democrats. OPM provides an agency on the low end of the scale and the Coast

Guard provides an agency on the high end of the scale. Lastly, OMB was rated by 153 Republicans and 206 Democrats, OPM was rated by 125 Republicans and 149 Democrats, and the Coast Guard was rated by 15 Republicans and 12 Democrats. In sum, all three agencies satisfy the first criteria, OMB and OPM satisfy the second, and OPM and Coast Guard satisfy the third.

**Table E1. Models of Performance Estimated with Democrat Performance Ratings and Republican Performance Ratings Separately**

| Agency Performance (-2.5, 2.5) | Democrats | Republicans | Democrats | Republicans |
|---|---|---|---|---|
| Months Vacant (0-42) | -0.027 | -0.026 | -0.007 | -0.006 |
| | [0.007]** | [0.009]** | [0.008] | [0.011] |
| *Agency Controls* | | | | |
| EOP (0,1) | -0.485 | 0.24 | -0.869 | -0.145 |
| | [0.106]** | [0.128]* | [0.134]** | [0.167] |
| Cabinet Department (0,1) | 0.724 | 0.915 | 0.449 | 0.583 |
| | [0.130]** | [0.119]** | [0.178]** | [0.135]** |
| Whole Department (0,1) | -1.239 | -0.793 | -0.653 | -0.12 |
| | [0.232]** | [0.262]** | [0.303]** | [0.323] |
| Workforce Skill -- Obama Administration | | | 0.519 | 0.627 |
| | | | [0.226]** | [0.182]** |
| Constant | 0.146 | -0.174 | -0.06 | -0.372 |
| | [0.068]** | [0.082]* | [0.075] | [0.102]** |
| $R^2$ | 0.35 | 0.25 | 0.54 | 0.45 |
| N | 39 | 39 | 37 | 37 |

Note: ** significant at the 0.05 level; * significant at the 0.10 level in two-tailed tests. Dependent variable is agency performance estimate based upon aggregated responses to question: "How would you rate the overall performance of the following agencies in carrying out their missions?" (1-5). Models estimated with Ordinary Least Squares and standard errors are clustered by department.

They models confirm what is reported in the main text. We note, however, that in models that also include controls for Obama-era workforce skills, the estimates get much smaller and imprecise. With the small N and the limited degrees of freedom, it is difficult to know how much weight to put on these results. We recommend readers apply appropriate caution in interpreting the results and recommend further research of this type.

# Appendix F. Collecting Vacancies and Turnover Data

To track vacancies and turnover, we started by identifying the relevant Senate confirmed positions (PAS). To identify the most proximate PAS position and higher level PAS position, we used the 2016 publication, *United States Government Policy and Supporting Positions* (i.e., the Plum Book).[18] In cases where the *Plum Book* was not clear, we used agency organization charts. We list below the organization charts we consulted:

- https://www.usda.gov/sites/default/files/documents/usda-organization-chart.pdf
- https://www.commerce.gov/sites/default/files/media/files/2015/docorgchartfinal.pdf
- https://fas.org/irp/agency/dod/org-man.pdf
- https://www2.ed.gov/about/offices/or/index.html
- https://www.energy.gov/leadership/organization-chart
- https://www.dhs.gov/sites/default/files/publications/19_1205_dhs-organizational-chart.pdf

*Vacancy Length*

We identify the first appointee's confirmation date using the *Washington Post* appointee tracker or senate.gov.[19] To calculate days until confirmation, we count the number of days from January 20, 2017 to the first confirmed appointee. If no appointee was confirmed, we counted days until June 22, 2020. If the first appointee left before the term was over, we determined the last date of service using the *Washington Post* appointee tracker or the appointee page on *Leadership Connect*. We then repeat this process of the second appointee left before the term was over. In total we collected data on the length of the vacancy at the start of the term and the vacancies that came after the first confirmation.

*Turnover*

---

[18] https://www.govinfo.gov/content/pkg/GPO-PLUMBOOK-2016/pdf/GPO-PLUMBOOK-2016.pdf
[19] https://www.washingtonpost.com/graphics/politics/trump-administration-appointee-tracker/database/

To determine the number of leaders/turnover, we used the quarterly editions of the *Federal Yellow Book*.[20] To begin, during the first quarter of 2017, we observed who held each position, whether that person was confirmed or acting or whether the *Federal Yellow Book* listed the position as vacant. We then tracked each subsequent quarter to see if there had been a change. If a position was vacant in the first quarter but filled in the second, we did not consider this turnover because it is not a shift from one person to another.[21] For board and commission chairs we used agency websites, either current or pages available through the *Wayback Machine* (archive.org). We begin by identifying the chair and term length, then track the chair through searches of websites through various dates of the Trump Administration, and then verify service length using *Leadership Connect*.[22]

---

[20] The *Federal Yellow Book* is a directory of federal government leaders. Leadership Directories, Inc. publishes this directory every quarter (https://www.leadershipconnect.io/products/print-leadership-directories/).

[21] It should be noted, however, that because we look at quarters, we may miss some turnover. For example, the Secretary of Commerce position was held by an acting official from January 20, 2017 to January 27, 2017, when Wilbur Ross was confirmed. Because we rely on quarterly data, we do not observe this transition.

[22] https://www.leadershipconnect.io/

**Appendix G. The Effect of Higher Level Vacancies on Federal Agency Performance**

Many federal agencies serve under more than one layer of Senate confirmed political appointee. For example, the Bureau of Land Management is headed by a Senate confirmed director. This director serves under another Senate confirmed appointee, the Assistant Secretary for Land and Minerals Management. The manuscript focuses on the most proximate vacancies to federal agencies, in the case of the BLM, the director. We might also be interested in the effects of vacancies in higher-level positions such as the Assistant Secretary for Land and Minerals Management.

In Table G1 we include estimates of models that include measures of vacancy length for both the most proximate PAS position (e.g., Director) and the next highest PAS position (e.g., Assistant Secretary) and their interaction (Model 2). The models are estimated on the 58 cases where an agency served under at least 2 levels of PAS appointees. This rules out all cabinet departments, agencies in the EOP, and independent agencies because in each case there is no higher PAS official. The small number of cases limits our flexibility in model specification.

The models reveal three interesting patterns. First, vacancies in the most proximate PAS positions are almost always correlated with lower perceptions of performance. These effects are most pronounced when there is a confirmed PAS appointee serving at a higher level. Second, holding constant vacancy length in the closest PAS position, longer vacancies in higher level PAS positions are estimated to *improve* agency performance. So, a bureau in DHS like the Federal Emergency Management Agency or Immigration and Customs Enforcement might operate slightly better without a confirmed DHS secretary. Third, long vacancies in both levels are better than a long vacancy in the most proximate PAS position but a short vacancy in the higher-level PAS position. The models provide some suggestive evidence that having bureaus run by higher level appointees due to a vacancy in the bureau is bad for performance and that agencies used to long vacancies may adjust and actually perform better in some cases. This is a fruitful avenue for future research.

**Table G1. OLS Models of Performance Estimates and PAS Vacancies at Different Levels of Hierarchy, 2020**

| Agency Performance (-2.5, 2.5) | (1) | (2) |
|---|---|---|
| Months Vacant (0-42) | -0.014 | -0.033 |
| | [0.006]** | [0.007]** |
| Leadership Transitions (0-4) | 0.106 | 0.104 |
| | [0.045]** | [0.038]** |
| Priority Department (0,1) | -0.174 | -0.179 |
| | [0.215] | [0.182] |
| Workforce Skill -- Obama Administration | 0.716 | 0.687 |
| | [0.086]** | [0.085]** |
| Agency Ideology (L-C) | -0.036 | -0.018 |
| | [0.113] | [0.114] |
| Months Vacant Higher Position (0-42) | 0.014 | -0.010 |
| | [0.005]** | [0.010] |
| Months Vacant*Months Vacant Higher Position | | 0.001 |
| | | [0.001]** |
| Constant | 0.195 | 0.568 |
| | [0.194] | [0.164]** |
| $R^2$ | 0.47 | 0.50 |
| N | 58 | 58 |

Note: *p<0.1; **p<0.05 in two-tailed tests. Dependent variable is agency performance estimate based upon aggregated responses to question: "How would you rate the overall performance of the following agencies in carrying out their missions?" (1-5). Models estimated with Ordinary Least Squares and standard errors are clustered by department.

## Appendix H. The Timing of Vacancies and Performance

To measure vacancies in the main specifications, we simply added months together based upon information in the quarterly editions of the *Federal Yellow Book*. This conflates the timing of a vacancy with the length of a vacancy. If a position has been vacant 6 months, for example, this could be at any point in the Trump Administration up to the time of the survey. However, if a position was vacant 40 months this implied a vacancy at the start, middle, and end of the term. To explore this further we estimate a series of models that measure vacancies in different ways. In the first set of models, we include the number of months a position was vacant at the start of the administration. In the second set of models, we include two measures-- the number of months a position was vacant at the start of the administration and the number of months the position was vacant after this first vacancy. In a final set of models we include indicators for whether a position was vacant or filled by a Senate confirmed appointee for each quarter between the Winter of 2017 and Summer 2020.

**Table H1. OLS Models of the Effect of Vacancies on Agency Performance Using Months to First Confirmed Appointee as Measure of Vacancy, 2020**

| Agency Performance (-2.5, 2.5) | All Agencies | All Agencies | PAS Head Only | PAS Head Only |
|---|---|---|---|---|
| Months to Confirm (0-42) | -0.009 | -0.024 | -0.009 | -0.029 |
| | [0.003]** | [0.010]** | [0.004]** | [0.009]** |
| Direct PAS Leader (0,1) | -0.133 | -0.132 | | |
| | [0.116] | [0.111] | | |
| Leadership Transitions (0-4) | 0.056 | 0.057 | 0.072 | 0.076 |
| | [0.031]* | [0.031]* | [0.040]* | [0.039]* |
| EOP (0,1) | -0.81 | -0.759 | -0.809 | -0.746 |
| | [0.080]** | [0.081]** | [0.082]** | [0.079]** |
| Cabinet Department (0,1) | 0.337 | 0.367 | 0.336 | 0.366 |
| | [0.073]** | [0.067]** | [0.077]** | [0.064]** |
| Office of the Secretary (0,1) | -0.484 | -0.56 | -0.486 | -0.577 |
| | [0.138]** | [0.142]** | [0.149]** | [0.144]** |
| Independent Commission (0,1) | -0.042 | 0.037 | -0.081 | 0.011 |
| | [0.089] | [0.100] | [0.095] | [0.095] |
| Priority Department (0,1) | -0.318 | -0.323 | -0.327 | -0.333 |
| | [0.147]** | [0.139]** | [0.163]* | [0.153]** |
| Workforce Skill -- Obama Administration | 0.584 | 0.571 | 0.641 | 0.631 |
| | [0.073]** | [0.073]** | [0.084]** | [0.083]** |
| Agency Ideology (L-C) | 0.051 | 0.043 | 0.025 | 0.014 |

|  |  |  |  |  |
|---|---|---|---|---|
|  | [0.077] | [0.080] | [0.092] | [0.094] |
| Months to Confirm^2 |  | 0.00 |  | 0.00 |
|  |  | [0.000] |  | [0.000]* |
| Constant | 0.263 | 0.325 | 0.122 | 0.198 |
|  | [0.153] | [0.132]** | [0.082] | [0.069]** |
| R² |  | 0.52 | 0.52 | 0.51 | 0.52 |
| N |  | 119 | 119 | 95 | 95 |

Note: * p<0.1; ** p<0.05. Dependent variable is agency performance estimate based upon aggregated responses to question: "How would you rate the overall performance of the following agencies in carrying out their missions?" (1-5). Models estimated with Ordinary Least Squares and standard errors are clustered by department. Models 3, 4 estimated using only agencies whose immediate head is a Senate-confirmed political appointee (PAS).

The models in Table H1 largely confirm the conclusions of the main text. Long vacancies at the start of a president's term due to slow confirmations is negatively correlated with our perceptual measure of performance. When asked about performance in the Summer of 2020, federal executives rated agencies lower that had slow confirmation processes. The curvilinear effect in these models is not quite so pronounced.

In Table H2 we combine these measures with measures of the number of other months a position was vacant after an initial confirmation. These models show a similar pattern. Vacancies at the start or vacancies at the end of a president's term are correlated with poorer perceptions of agency performance with the later vacancies having a more pronounced effect on perceptions of performance.

**Table H2. OLS Models of the Effect of Vacancies on Agency Performance Using Months to First Confirmed Appointee and Other Months Vacant as Measures of Vacancy, 2020**

| Agency Performance (-2.5, 2.5) | All Agencies | All Agencies | PAS Head Only | PAS Head Only |
|---|---|---|---|---|
| Months to Confirm (0-42) | -0.012 | -0.037 | -0.016 | -0.047 |
|  | [0.003]** | [0.011]** | [0.005]** | [0.009]** |
| Vacancy Other than 1st Vacancy (0-28) | -0.02 | -0.061 | -0.036 | -0.082 |
|  | [0.011]* | [0.032]* | [0.016]** | [0.032]** |
| Direct PAS Leader (0,1) | -0.203 | -0.167 |  |  |
|  | [0.137] | [0.130] |  |  |
| Leadership Transitions (0-4) | 0.112 | 0.13 | 0.149 | 0.174 |
|  | [0.050]** | [0.051]** | [0.059]** | [0.060]** |
| EOP (0,1) | -0.782 | -0.733 | -0.784 | -0.712 |
|  | [0.090]** | [0.084]** | [0.096]** | [0.079]** |
| Cabinet Department (0,1) | 0.345 | 0.352 | 0.347 | 0.356 |
|  | [0.071]** | [0.070]** | [0.080]** | [0.075]** |

| | | | | |
|---|---|---|---|---|
| Office of the Secretary (0,1) | -0.543 | -0.686 | -0.6 | -0.762 |
| | [0.131]** | [0.137]** | [0.154]** | [0.143]** |
| Independent Commission (0,1) | -0.074 | 0.005 | -0.126 | -0.027 |
| | [0.099] | [0.104] | [0.101] | [0.101] |
| Priority Department (0,1) | -0.314 | -0.293 | -0.289 | -0.285 |
| | [0.144]** | [0.128]** | [0.161]* | [0.137]* |
| Workforce Skill -- Obama Administration | 0.577 | 0.538 | 0.612 | 0.582 |
| | [0.076]** | [0.077]** | [0.080]** | [0.078]** |
| Agency Ideology (L-C) | 0.051 | 0.045 | 0.013 | 0.01 |
| | [0.073] | [0.079] | [0.089] | [0.092] |
| Months to Confirm^2 | | 0.001 | | 0.001 |
| | | [0.000]** | | [0.000]** |
| Vacancy Other than 1st Vacancy^2 | | 0.002 | | 0.002 |
| | | [0.001] | | [0.001]* |
| Constant | 0.344 | 0.481 | 0.163 | 0.364 |
| | [0.167]* | [0.153]** | [0.076]** | [0.078]** |
| R² | 0.53 | 0.55 | 0.55 | 0.57 |
| N | 119 | 119 | 95 | 95 |

Note: * p<0.1; ** p<0.05. Dependent variable is agency performance estimate based upon aggregated responses to question: "How would you rate the overall performance of the following agencies in carrying out their missions?" (1-5). Models estimated with Ordinary Least Squares and standard errors are clustered by department. Models 3, 4 estimated using only agencies whose immediate head is a Senate-confirmed political appointee (PAS).

In H3 we estimate models with indicators for vacancies in specific quarters. These estimates are largely inconclusive, suggesting that we cannot reject the null that vacancies in any one quarter have no effect on performance. A natural conclusion is that it is the accumulation of vacancy duration that influences perceptions of performance. The only exception to this general pattern is that vacancies at the time of the survey are negatively correlated with perceptions of performance. A vacant quarter is estimated to decrease an agency's rating by about 0.3, a little less than ½ a standard deviation. One conclusion, consistent with results from Table 2 is that earlier vacancies matter less than vacancies in the current quarter.

**Table H3. OLS Models of the Effect of Vacancies on Agency Performance Using Indicators of Vacant Quarters to Measure Vacancies, 2017-2020**

| Agency Performance (-2.5, 2.5) | All Agencies | All Agencies | All Agencies | All Agencies | All Agencies |
|---|---|---|---|---|---|
| Winter 2017 | -0.054 | -0.129 | | | |
| | [0.109] | [0.103] | | | |
| Spring 2017 | 0.088 | 0.073 | | | |
| | [0.183] | [0.173] | | | |
| Summer 2017 | -0.12 | -0.111 | | | |
| | [0.244] | [0.191] | | | |
| Fall 2017 | 0.023 | -0.133 | | | |

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| | [0.148] | [0.067]* | | | |
| Winter 2018 | -0.189 | | -0.275 | | |
| | [0.172] | | [0.153]* | | |
| Spring 2018 | -0.089 | | -0.054 | | |
| | [0.211] | | [0.250] | | |
| Summer 2018 | 0.139 | | 0.176 | | |
| | [0.273] | | [0.284] | | |
| Fall 2018 | 0.163 | | -0.034 | | |
| | [0.304] | | [0.331] | | |
| Winter 2019 | -0.239 | | | -0.185 | |
| | [0.316] | | | [0.235] | |
| Spring 2019 | 0.112 | | | 0.061 | |
| | [0.207] | | | [0.172] | |
| Summer 2019 | 0.062 | | | 0.158 | |
| | [0.380] | | | [0.321] | |
| Fall 2019 | -0.098 | | | -0.281 | |
| | [0.237] | | | [0.179] | |
| Winter 2020 | -0.057 | | | | -0.105 |
| | [0.136] | | | | [0.131] |
| Spring 2020 | 0.089 | | | | 0.068 |
| | [0.151] | | | | [0.116] |
| Summer 2020 | -0.342 | | | | -0.346 |
| | [0.109]** | | | | [0.112]** |
| Direct PAS Leader (0,1) | -0.113 | -0.149 | -0.098 | -0.182 | -0.151 |
| | [0.112] | [0.107] | [0.127] | [0.128] | [0.105] |
| Leadership Transitions (0-4) | 0.078 | 0.056 | 0.044 | 0.089 | 0.088 |
| | [0.040]* | [0.034] | [0.033] | [0.031]** | [0.033]** |
| EOP (0,1) | -0.849 | -0.828 | -0.876 | -0.819 | -0.84 |
| | [0.146]** | [0.125]** | [0.102]** | [0.107]** | [0.085]** |
| Cabinet Department (0,1) | 0.257 | 0.288 | 0.274 | 0.245 | 0.266 |
| | [0.094]** | [0.103]** | [0.077]** | [0.068]** | [0.071]** |
| Office of the Secretary (0,1) | -0.478 | -0.472 | -0.465 | -0.376 | -0.431 |
| | [0.177]** | [0.180]** | [0.129]** | [0.108]** | [0.106]** |
| Independent Commission (0,1) | 0.087 | -0.067 | -0.086 | -0.071 | 0.065 |
| | [0.196] | [0.105] | [0.109] | [0.122] | [0.133] |
| Priority Department (0,1) | -0.248 | -0.261 | -0.261 | -0.296 | -0.264 |
| | [0.155] | [0.158] | [0.156] | [0.134]** | [0.149]* |
| Workforce Skill -- Obama Administration | 0.552 | 0.548 | 0.551 | 0.566 | 0.572 |
| | [0.073]** | [0.063]** | [0.060]** | [0.064]** | [0.056]** |
| Agency Ideology (L-C) | 0.058 | 0.054 | 0.064 | 0.06 | 0.071 |
| | [0.079] | [0.079] | [0.083] | [0.068] | [0.065] |
| Constant | 0.296 | 0.355 | 0.227 | 0.257 | 0.217 |
| | [0.167]* | [0.185]* | [0.184] | [0.165] | [0.137] |
| $R^2$ | 0.52 | 0.47 | 0.48 | 0.48 | 0.5 |
| N | 125 | 125 | 125 | 125 | 125 |

Note: Note: * $p<0.1$; ** $p<0.05$. Dependent variable is agency performance estimate based upon aggregated responses to question: "How would you rate the overall performance of the following agencies in carrying out their missions?" (1-5). Models estimated with Ordinary Least Squares and standard errors are clustered by department. Models 3, 4 estimated using only agencies whose immediate head is a Senate-confirmed political appointee (PAS).

**Appendix I. External Validation of Agency Performance Measures Using GAO Open Recommendations Data**

There is an important concern that subjective measures like the ones we generate can be quite removed from actual performance. To validate the measure we looked for objective measures of performance. This is difficult since few comparable objective measures exist. Indeed, the lack of such measures is one reason why we pursued this project. Some comparable objective measures focus on ancillary tasks (e.g., procurement). Others focus on the accomplishment of goals, but the goals themselves vary in ambition and some are closer to core performance than others.

Ultimately, we focus on contemporaneous Government Accountability Office (GAO) evaluations of agencies. Since 2018, GAO has maintained a recommendations database that includes every GAO recommendation that remains open in response to an agency study or audit. The database signifies which recommendations they consider "priority" recommendations. GAO conducts hundreds of studies a year and the database includes the results of studies of more than 100 agencies.[23] The data generating process for this measure (and other objective measures) creates real complexity. Notably, large agencies are more likely to have a program audited by GAO because they have more programs. For example, GAO is likely to conduct more investigations and have more recommendations for the Department of Defense (DOD; 700,000 employees) than the Administrative Conference of the United States (10 employees) because DOD has dramatically more programs and comprises many more agencies. This makes disentangling performance from size complicated. Agencies may also not implement recommendations because they disagree with GAO's assessment. GAO reports that 75% of their recommendations are implemented. Still, this is GAO's best information about agency struggles, enduring and new at the end of 2020.

---

[23] https://www.gao.gov/reports-testimonies/recommendations-database.

With these caveats, we regress the number of open and priority recommendations at the end of December, 2020 on our performance estimate for the agency, controlling for the log of agency size. We also estimate the models excluding the performance scores for agencies where Republicans and Democrats disagreed when we estimated scores separately. We repeat our analysis with 2021 data since our performance estimate in 2020 should also predict performance in 2021. Our aim is to determine whether there is a correlation between our measure and the objective measure, once we account for size, rather than specify a full model of GAO recommendations.

We report the results of these models for 2020 and 2021 in Tables I1 and I2, respectively. Notably, higher performance estimates are correlated with fewer open recommendations in the models. The results are strongest in 2020 and for counts of all recommendations, perhaps because GAO finds fewer priority fixes in smaller agencies. Such agencies have fewer broadly impactful programs. A one unit increase in an agency's performance estimate correlates with a 1-2 recommendation decrease in the predicted number of open priority recommendations (with an 4-5 agency yearly average) and a 10 to 20 reduction in all recommendations (with a 34 to 52 agency yearly average).

It is important to remember that there are other potential objective measures, including accomplishment of various agency performance goals, employee awards, inclusion of programs on the GAO's high risk list. Some of these measures are a closer measure of performance than others. For example, employee awards are a bit further from agency performance. That data generating process for each is different and presents different challenges. For example, we do not know whether agencies have chosen hard or easy goals and it is unclear whether agencies with no programs on the high risk list are omitted because they were performing well or because they were not evaluated. So, the evidence here should be considered confirmatory but not definitive along with the comparisons to subjective measures in the text.

**Table I1. Negative Binomial Regression Models of Open GAO Recommendations by Agency, 2020**

| | Priority Rec. | All Rec. | Priority Rec. | All Rec. |
|---|---|---|---|---|
| | All Cases | All Cases | Exclude Cases w/ Partisan Disagreement | Exclude Cases w/Partisan Disagreement |
| Performance Rating | -0.578 | -0.634 | -0.529 | -0.553 |
| | [0.253]** | [0.233]** | [0.252]** | [0.235]** |
| Ln(Employment in 1,000s) | 0.414 | 0.513 | 0.277 | 0.4 |
| | [0.088]** | [0.064]** | [0.093]** | [0.066]** |
| Constant | 0.471 | 2.394 | 0.564 | 2.472 |
| | [0.349] | [0.260]** | [0.312]* | [0.253]** |
| Ln(α) | 1.078 | 0.307 | 1.127 | 0.278 |
| | [0.195]** | [0.115]** | [0.221]** | [0.118]** |
| N | 90 | 90 | 82 | 82 |

Note: * p<0.1; ** p<0.05. Dependent variable is count of number of open GAO recommendations in 2020. Models 1 and 3 include only recommendations GAO considers priority recommendations and Models 2 and 4 include all recommendations. Models 3 and 4 exclude ratings where Republican and Democratic-only ratings were significantly different.

**Table I2. Negative Binomial Regression Models of Open GAO Recommendations by Agency, 2021**

| | Priority Rec. | All Rec. | Priority Rec. | All Rec. |
|---|---|---|---|---|
| | All Cases | All Cases | Exclude Cases w/ Partisan Disagreement | Exclude Cases w/Partisan Disagreement |
| Performance Rating | -0.464 | -0.522 | -0.447 | -0.468 |
| | [0.243]* | [0.230]** | [0.245]* | [0.238]** |
| Ln(Employment in 1,000s) | 0.299 | 0.446 | 0.223 | 0.384 |
| | [0.081]** | [0.076]** | [0.080]** | [0.076]** |
| Constant | 0.631 | 2.286 | 0.692 | 2.343 |
| | [0.324]* | [0.305]** | [0.307]** | [0.302]** |
| Ln(α) | 0.972 | 0.293 | 1.073 | 0.328 |
| | [0.197]** | [0.142]** | [0.203]** | [0.135]** |
| N | 90 | 90 | 82 | 82 |

Note: * p<0.1; ** p<0.05. Dependent variable is count of number of open GAO recommendations in 2020. Models 1 and 3 include only recommendations GAO considers priority recommendations and Models 2 and 4 include all recommendations. Models 3 and 4 exclude ratings where Republican and Democratic-only ratings were significantly different.