Niskanen Center

Scott Kupor, Director
Office of Personnel Management
1900 E St NW
Washington, D.C., 20415

March 26, 2026

Re: RIN 3206-AO88 "Managing Senior Professional Performance" [Docket ID: OPM-2025-0011] & RIN 3206-AP06 "Performance Appraisal for General Schedule, Prevailing Rate, and Certain Other Employees" [Docket ID: OPM-2025-0273]

Dear Director Kupor,

On behalf of the Niskanen Center, I am pleased to share comments on the Office of Personnel Management's (OPM) proposed rules regarding "Managing Senior Professional Performance" and "Performance Appraisal for General Schedule, Prevailing Rate, and Certain Other Employees" published in the Federal Register on February 24, 2026. Enclosed along with this letter is a lengthier, detailed description (pages 5-17) of our observations and recommendations for OPM to consider in revising its approach to this topic.

Background
The Niskanen Center is a nonprofit public policy organization that advocates for a government that provides social insurance and essential public goods, fosters market competition and innovation, invests in state capacity, and does not impede productive enterprise. We are committed to the principles of liberal democracy and an open society that encourages engagement, cooperation, discussion, and learning.

An agile, accountable, and high-performing public workforce is essential to our vision of a stronger government and better governing outcomes for all Americans. Our reform agenda for the federal workforce revolves around several core design principles:

- **Fewer, Simpler Personnel Systems** - Reduce the complexity of federal personnel systems to make them easier to understand, simpler to administer, and more flexible to accommodate changing conditions.
- **Fast & Fair Accountability** - Streamline and speed up the process for removing poor performers or those who engage in misconduct while maintaining Constitutional guardrails against arbitrary partisan dismissals and unlawful discrimination.
- **Market-Sensitivity** - Prioritize the federal government's ability to compete in the labor market in approaches to compensation, hiring, and benefits administration so that the federal government can attract and retain the best talent on offer.
- **Minimized HR Touches, Maximized Manager Flexibility** - Devolve as much authority to agencies and then to line managers as possible, recognizing that centralized HR

functions have no accountability for mission outcomes and minimal ability to weigh tradeoffs in their pursuit.

- **Internal Capacity First** - Invest first in internal agency capacity before looking to augment agency workforces with vendors who may have misaligned incentives and are more difficult to manage in the long term, especially for core agency work not otherwise happening in the private sector.

Underpinning each of these principles is also a commitment to uphold **nonpartisanship and observable merit** as the foundation for all policy choices related to the federal workforce–without a bedrock commitment to a merit system free from partisan meddling, no meaningful reform is possible that addresses the actual root causes of state incapacity.

Accordingly, we believe OPM's goal with both rulemakings (i.e., to improve the adoption and administration of agency performance management systems such that they better recognize relative differences in demonstrated performance) is laudable. In particular, we think OPM's move to remove "Level 2" rating is a step in the right direction. This change will help clarify for supervisors and staff which performance is *actually* unacceptable and provide both sides with clarity about where individual employees stand. This change should have been made a long time ago. Similarly, we support requirements for OPM to continuously certify performance management systems to continually raise the bar for agencies.

OPM's rule misstates understates the evidence and elides instructive history

However, we would also be remiss if we didn't point out that OPM's intended strategy for improving the rating system—employing forced ranking distributions for federal employees—seems unlikely to achieve a favorable outcome. We think OPM needs to be more about how hard this task is, the limits of the evidence it relies on, and the genuinely mixed experience generations of employers have had with employing strategies like forced ranking to do so. Intellectual rigor and regulatory honesty demand a fuller accounting of the facts than is present in the NPRMs today.

For example, OPM cites an experimental study [1] and a literature review [2] to justify the potential benefits of forced rating distribution. In the former case, OPM is correct that the study shows some potential benefits to implementing such a policy (albeit in a laboratory setting) but does not adequately consider the authors' other conclusions, including that the strategy may have detrimental long-term impacts on performance. This includes a suggestion that "it may be

[1] Johannes Berger, et al., "Performance Appraisals and the Impact of Forced Distribution: An Experimental Investigation." IZA Discussion Paper, No. 5020 (2010).

[2] Aniek Wijayanti et al., "What Do We Know about the Forced Distribution System: A Systematic Literature Review and Opportunities for Future Research," *Management Review Quarterly* 75, no. 1 (2025): 747–88, https://doi.org/10.1007/s11301-023-00396-8

problematic to set up a forced distribution when employees have experienced a more 'liberal' system of performance evaluations before."[3] Given OPM's other statements about the current system, we presume that you would agree that "a more liberal system" likely characterizes the current state of federal performance management. In the case of the literature review, the authors discuss at length how mixed the literature is on the topic and one of their conclusions is that "[t]he risks associated with [forced distribution] tend to outweigh the benefits when applied to jobs involving task interdependence"[4] which also surely describes many jobs within the highly complex federal environment.

Further, the instances OPM cites of companies that have adopted forced ranking have not uniformly had a good experience. Microsoft[5] and General Electric,[6] for example, very famously *walked away* from their forced rating distribution schemes after long periods of pushback from management and employees. While some companies, including and especially in the technology industry, continue to use such systems, the evidence that it is "working" is thin at best. And, in any case, these employers are significantly more homogenous than the federal government in terms of position, labor market, and mission diversity, and have other personnel flexibilities (e.g., on compensation, variable compensation, etc.) not present in the government. It isn't clear that even a system with overwhelming positive results in those industries would port neatly into the federal government's old and badly-in-need-of-reform personnel system.

Finally, we think OPM should take more seriously the government's own history with such systems. This is not the first time that the government has sought to address these challenges. In some sense, this is one of the oldest and most intractable problems behind the basic question of classification and compensation. It would be worthwhile for OPM to more squarely grapple with why those schemes didn't work and what might be different this time.

None of this is to suggest that reform isn't necessary or that forced ranking could not work in some part of the government. Rather, we believe that the rule's detractors, supporters, and OPM itself all need to acknowledge how little we really know about what will "work" in the federal context. The antibodies come out in both directions because everyone is simply speculating about what the impact of changes will be.

---

[3] *Berger, et al.* 30-31.
[4] *Wijayanti et al.* 782
[5] Tom Warren, "Microsoft Axes Its Controversial Employee-Ranking System," *The Verge*, November 12, 2013, https://www.theverge.com/2013/11/12/5094864/microsoft-kills-stack-ranking-internal-structure.
[6] Max Nisen, "Why GE Had to Kill Its Annual Performance Reviews after More than Three Decades," Economic Indicators, *Quartz*, July 20, 2022, https://qz.com/428813/ge-performance-review-strategy-shift.

OPM should employ its statutory demonstration authority to generate evidence
To that end, we propose that OPM modifies its approach to answer some of these questions before proceeding with broader implementation. **OPM should use this as an opportunity to launch several pilots using its statutory demonstration project authority to test whether and where forced ranking (and other more novel schemes) produces better governing outcomes.** In an ideal world, OPM would use randomization, rigorous evaluation by independent experts, and other strategies that produce reliable evidence upon which policymakers could rely in considering a broader policy change.

This approach is the strongest defense against the charge (which, while we do not agree, we expect to be the thrust of many other comments) that the proposed reform motivated by OPM is motivated by partisan concerns rather than genuine interest in workforce improvement. If forced ranking works, a rigorous evaluation will demonstrate that it works, and the case for government-wide adoption will be considerably stronger than any theoretical argument OPM can offer *ex ante*. If it does not work (or if it works in some contexts but not others) the evaluation will reveal that too, and the government will have avoided imposing a policy on its entire workforce that has negative impacts on performance or morale.

More fundamentally, this approach reflects a commitment to governing by evidence rather than by assumption. The federal workforce is too important and the missions it carries out are too consequential to subject to large-scale experiments without control, without independent evaluation, and without a genuine willingness to learn from the results. OPM has the authority and the institutional capacity to do this right. We urge it to take that path, having the courage to ask hard questions and seek answers in ways that it has not historically been willing to do during administrations of either party.

Regardless of OPM's final regulation on this topic, however, ultimate responsibility to structure the federal personnel system rests with Congress and we continue to believe that durable reform is only possible through Congressional action. In parallel, we continue to advocate for the Executive Branch to thoughtfully engage their counterparts on the Hill and in civil society to re-think the performance management and accountability processes. We would welcome the opportunity to work with all interested parties on such an effort.

Sincerely,

Gabe Menchaca
Senior Policy Analyst
Niskanen Center

* * * * *

Niskanen Center

Additional Detail

Performance management is a timeless problem and few employers, in any sector, are very good at it

We begin with a premise that OPM's proposed rule takes largely for granted: that the challenge of accurately evaluating employee performance is, at bottom, a regulatory design problem that the right set of rules can solve. This premise deserves more scrutiny than OPM gives it. The history of performance management — across sectors, across decades, across every conceivable organizational structure — suggests something more humbling. Nearly every large employer that has attempted to systematically evaluate relative employee performance has struggled, iterated, and often retreated. The federal government is not uniquely dysfunctional in this regard. It is navigating a problem that has defied elegant solutions as long as employers have had sufficient complexity to require impersonal management systems.

The basic architecture of modern formalized performance ratings traces back roughly a century or four depending on when you start the clock. The early Jesuit Order, for instance, kept detailed assessments of individual personnel performance (*catalogi* and *informationes*) that reflected the complexity involved in administering such a far-flung organization and which they arguably adapted from other organizations around them. [7] A couple hundred years later, a need to evaluate and deploy military officers led American military commanders in the aftermath of the Revolutionary War and the War of 1812 to similarly come up with rudimentary schemes for soliciting performance information about various parts of the officer corps. [8]

More recognizable modern or "scientific" approaches started to appear during the early 20th century, originating from the same impulse that animated Fredrick Taylor and Henry Ford. The private sector adopted similar frameworks in the postwar period as firms grew large enough that managers could no longer rely on personal knowledge of every employee. [9] By the 1950s, performance appraisal had become a standard feature of corporate human resources. And from the very beginning, the same challenges were already present that drive today's debate:

---

[7] Markus Friedrich, "A Jesuit Culture of Records?: The Society of Jesus, the Life Cycle of Administrative Documents, and the Late Medieval and Early Modern History of Bureaucratic Information," *The International Symposia on Jesuit Studies* 1, no. 1 (2021), https://doi.org/10.51238/ISJS.2019.06.

[8] Malin Craig Jr., *History of the Officer Efficiency Report System, United States Army 1775-1917* (Office of the Chief of Military History, 1953).

[9] Peter Cappelli and Anna Tavis, "The Performance Management Revolution," Employee Performance Management, *Harvard Business Review*, October 1, 2016, https://hbr.org/2016/10/the-performance-management-revolution.

ratings clustered at the top of the scale, supervisors avoided delivering critical feedback, schemes were inhumane and discriminatory, and the systems failed to accurately distinguish between good and mediocre performance. [10]

These are not uniquely governmental failures; they are features of human psychology operating within organizational incentive structures. Leniency error, for example, has been documented for several decades as a challenge in designing subjective rating scales. [11] Supervisors in any setting face a basic tension: the person whose performance they are evaluating is also someone they must work with, motivate, and depend upon. Delivering honest critical feedback carries real or perceived interpersonal costs, and in most organizations, there is often no countervailing incentive strong enough to consistently overcome the resulting natural impulse toward grade inflation. This dynamic has been documented extensively in organizational management literature. [12] OPM's proposed rule acknowledges the phenomenon but understates its universality.

The private sector's experience with this problem is instructive precisely because it demonstrates how persistent the challenge is even with vastly more managerial flexibility and financial incentive power than the federal government possesses. Large white-collar employers have cycled through annual reviews, continuous feedback models, 360-degree assessments, calibration committees, and various rating scales—and the fundamental complaints have remained remarkably stable. [13] Some employers with highly transactional work that lends itself to objective measurement—call centers tracking handle times, sales teams tracking quota attainment, warehouse operations tracking throughput—have achieved more defensible differentiation through automatic tracking. Other employers have tried to graft the two strategies by conducting more automated tracking of employee activity (e.g., keystrokes, eye movement, geolocation, etc.) but even in those contexts there is disagreement about the efficacy of such strategies. In 2024, for example, GAO published a report on "digital

[10] Douglas McGregor, "An Uneasy Look at Performance Appraisal," Managing People, *Harvard Business Review*, republished September 1, 1972, originally published May-June 1957, https://hbr.org/1972/09/an-uneasy-look-at-performance-appraisal.
[11] J. P. Guilford, *Psychometric Methods*, 2nd ed. (McGraw-Hill, 1954).
[12] Frank J. Landy and James L. Farr, "Performance Rating," *Psychological Bulletin* (US) 87, no. 1 (1980): 72–107, https://doi.org/10.1037/0033-2909.87.1.72.
[13] Derek Thompson, "The Case Against Performance Reviews," Business, *The Atlantic*, January 29, 2014, https://www.theatlantic.com/business/archive/2014/01/the-case-against-performance-reviews/283402/.

surveillance" strategies that found genuinely mixed views across industry, labor, and vendors, and others about the prevalence and utility of these solutions. [14]

The basic problems at the heart of any performance management system are well known and worth naming plainly, because OPM's proposed rule implicitly assumes that most of them can be solved by changing the distribution of ratings:

- **Supervisor quality varies enormously, and performance ratings are fundamentally acts of judgment.** The accuracy of any rating depends on the quality of the person making it. No rating architecture can compensate for a supervisor who lacks management skill, institutional support, or general moral courage to differentiate honestly. This is not a problem that a standardized distribution solves; it is a problem that a standardized distribution can worsen if it forces inexperienced or poorly trained supervisors to make finer distinctions than they are equipped to draw.
- **Ratings cluster at the upper end of virtually every scale ever devised.** This is not a uniquely federal pathology. It occurs in Fortune 500 companies, in universities, in the military, in nonprofits, and in small businesses. The tendency is psychological and structural because the costs of giving a low rating (interpersonal conflict, potential grievance, loss of employee motivation) are borne directly and immediately by the supervisor, while the benefits of honest differentiation (better organizational resource allocation, improved aggregate performance) are diffuse and accrue to the firm more broadly.
- **Organizations and individuals are broadly averse to giving bad feedback.** This is related to, but distinct from, leniency bias in ratings. Even supervisors who rate accurately may fail to communicate the substance of their assessment clearly enough for the employee to act on it. The feedback problem is at least as consequential as the rating problem, and a standardized distribution does nothing to address it directly.
- **The administrative burden of performance management competes with the work itself.** For example: it would be valuable if every supervisor produced a detailed written assessment of each direct report every week. It would also be enormously time-consuming, and the time spent on documentation is not spent on mission delivery. Every performance management system must navigate this tradeoff, and the systems that demand the most rigor from supervisors tend to generate the most resistance and the most perfunctory compliance.

[14] *Digital Surveillance of Workers: Tools, Uses, and Stakeholder Perspectives*, Q&A Report to Congressional Requesters GAO-24-107639 (Government Accountability Office, 2024), https://www.gao.gov/assets/gao-24-107639.pdf.

Private sector employers have addressed these challenges in a variety of ways—mandating frequent informal feedback sessions, experimenting with light-touch rating systems, [15] investing significant resources in calibration processes to create norms across business units, [16] defaulting where possible to metrics that reduce the role of subjective judgment, and (yes) forced ranking distribution. But none of these approaches has proven universally effective, and the fact that the private sector—or even companies within a single industry—has not converged on a single model should tell us something important: this is an extraordinarily hard problem that does not lend itself to simple heuristic answers. The proposed rule's implicit assumption that the private sector has discovered a solution in forced ranking that the federal government has merely been too timid to adopt is not supported by the weight of the evidence.

The operating environment for the federal government makes this problem harder, not easier

If performance management is difficult in the private sector—where employers have broad authority to set compensation, restructure teams, and terminate employees with relative speed—it is substantially more difficult in the federal government, where additional layers of legal, structural, and political complexity constrain every dimension of the problem. OPM's proposed rule does not adequately reckon with these constraints, and several of them bear directly on whether a standardized distribution of ratings is likely to achieve its stated objectives:

- **The management skills of federal supervisors are, on average, unusually low relative to the complexity of what they are being asked to do.** This is not an insult; it is a structural observation about selection incentives. On the career side, the rigid classification structure of the General Schedule means that the primary route to meaningful salary growth for high-performing individual contributors is to move into management—regardless of whether they have any aptitude or interest in managing people. The GS system effectively selects for technical expertise and willingness to accept supervisory responsibility, not for management skill. Furthermore, unlike in much of the private sector, management training for career employees making this transition is chronically under-resourced. OPM's own proposed rule acknowledges this by requiring a new supervisory critical element, but it is not clear that this alone will be enough to overcome structural barriers to good supervision.

---

[15] Marcus Buckingham and Ashley Goodall, "Reinventing Performance Management," Employee Performance Management, *Harvard Business Review*, April 1, 2015, https://hbr.org/2015/04/reinventing-performance-management.
[16] "What HR Leaders Need to Know About Performance Calibration," Korn Ferry, accessed March 25, 2026, https://www.kornferry.com/insights/featured-topics/employee-experience/hr-leaders-and-performance-calibration.

On the political side, the need for tight partisan alignment with the President means that many management roles are occupied by political appointees—activists, think tank alumni, and former Congressional staff—who are often managing organizations at scale for the first time in their careers and may themselves be poorly trained in the mechanics of management. The high turnover of political appointees, and the absolute turnover that accompanies every presidential transition, means these leaders are frequently not incentivized to invest in the long-term health of the organizations they manage. A standardized distribution imposed by OPM and administered in practice by this population of supervisors is a system whose accuracy depends entirely on the judgment of people who, structurally, are the least well-positioned to exercise that judgment well.

- **The level of scrutiny on both favorable and unfavorable performance ratings is significantly higher in the federal context** because both public benefits and individual constitutional rights are implicated. Federal employment carries a property interest protected by the Fifth Amendment's due process guarantee. [17] As a result, public performance management must satisfy procedural and substantive constraints that have no private-sector analogue. This is a dimension of complexity that transcends even the application of equal employment opportunity law—which itself applies more stringently to the federal government as both a model employer and an employer with an unusually bright spotlight cast on it by the media and opposition parties. Performance management in the federal sector is embedded within a constitutionally protected framework that makes the stakes of inaccuracy higher and the consequences of arbitrary action more severe.

- **The goals of federal work are often inscrutable, conflicting, and difficult to measure on any annual cycle.** The work of many federal agencies is abstract, symbolic, or measured on extraordinarily long timescales. It is genuinely difficult to construct annual performance standards based on objective criteria when the underlying work does not produce easily quantifiable outputs. This is, in part, why the Government Performance and Results Act (GPRA) [18] and its successor GPRAMA [19] have had uneven success: it sounds compelling to develop performance measures, manage to them, and publish

[17] Evan Chinn and Chloe Lee, "Understanding Loudermill Rights: Balancing Due Process in Employee Discipline and Best Practices," Municipal Research and Services Center of Washington (MRSC), August 5, 2024, https://mrsc.org/stay-informed/mrsc-insight/august-2024/loudermill-rights.
[18] Government Performance and Results Act of 1993, Pub. L. No. 103–62, 107 Stat 285 (1993), https://www.govinfo.gov/content/pkg/STATUTE-107/pdf/STATUTE-107-Pg285.pdf.
[19] GPRA Modernization Act of 2010, Pub. L. No. 111–352 (2011), https://www.congress.gov/111/plaws/publ352/PLAW-111publ352.pdf.

progress, but it has proven extremely hard to do this at scale in the government. The Social Security Administration has many obvious metrics—call center wait times, improper payment rates, processing backlogs, etc.—that lend themselves to objective performance standards during a single rating period. The Department of Energy's Office of Science does not. A standardized distribution of ratings that works reasonably well in an agency with measurable outputs may produce deeply arbitrary results in an agency where the work resists quantification or where the time to results expands beyond a single rating cycle.

- **The dual accountability that federal employees owe to both Congress and the President creates a structural tension that complicates any unified performance management framework.** Federal employees serve, in a meaningful sense, two bosses who may not agree about anything or everything: Congress, which created their positions and authorized their programs for specified purposes, and the President, whose subordinates hire them, direct their work and rate their performance. When Congress and the President disagree about an agency's mission or priorities, neutral administration of the personnel rules can be caught in the middle. Congress has historically responded to this tension by legislating constraints on processes, which is its prerogative under the Constitution, but their intervention has a way of still distorting behavior after the crisis has passed. A standardized distribution imposed by OPM cannot resolve this structural tension, and it may exacerbate it by giving Congress additional reason to intervene.

None of this is to suggest that the federal government should abandon the effort to improve performance management. It is to say that the difficulty of the problem is significantly greater than OPM's proposed rule acknowledges, and that the proposed solutions must be evaluated against this more realistic baseline. A system that works in theory but cannot be administered competently by the supervisory population that we have today, or that produces arbitrary results in agencies where work outputs resist measurement, is just a different kind of failure.

The federal government has been searching for 150 years to find a workable system

To that end, we certainly agree with OPM's contention that "Federal agencies have confronted persistent challenges in effectively and consistently managing employee performance." [20] As OPM notes, this has been a long-standing challenge that goes back further than OPM highlights in its proposed rule. Over the last 100+ years, Federal employees have been subject to a

[20] Personnel Management Office, "Performance Appraisal for General Schedule, Prevailing Rate, and Certain Other Employees," Federal Register, February 24, 2026, https://www.federalregister.gov/documents/2026/02/24/2026-03619/performance-appraisal-for-general-schedule-prevailing-rate-and-certain-other-employees#p-38.

variety of different schemes for performance ratings. Several efforts, for example, predate the more modern ones highlighted in OPM's preamble:

- During the late 1880s through the early 1900s, successive presidents from Benjamin Harrison through Theodore Roosevelt sought to establish some kind of efficiency rating system for employees. For Roosevelt, this became situated within a broader effort to modernize government management under the umbrella of the Keep Commission.[21]
- Coming out of these more piecemeal efforts, Congress directed the creation of a "Division of Efficiency" within the Civil Service Commission (CSC) in 1913 for the explicit purpose of designing systems to "ascertain[] how far each employee's work is above or below normal" and to connect those ratings with salary.[22] The Division, which became the independent Bureau of Efficiency in 1916, set about working with agencies to come up with a variety of such systems.
- The Classification Act of 1923 provided for the expansion of this system to covered employees, requiring that "[t]he head of each department shall rate in accordance with such systems the efficiency of each employee under his control or direction."[23] Various further modifications happened over the ensuing ~25 years, including the addition of language in 1941 that specified which ratings were required (i.e., "good or better than good") to get within-grade pay raises.[24]
- A few years later, Congress established a more modern system of performance ratings with the Performance Rating Act of 1950[25] which further refined the policy for performance rating by requiring approval of agency plans by the CSC, required warnings for unsatisfactory performance, establishing a right for impartial appeal, and clarified the Civil Service Commission's role in administering the system.

In other words, the federal government has sought a performance management system that accurately distinguishes relative employee quality for a long time, and the results have been uneven. In a sense, OPM's proposed rules here are merely the latest in a long line of attempts to get this right. There are reasons to be skeptical that this move to attempt forced ranking will "work" but that's not a reason to give up experimenting.

---

[21] W. Brooke Graves et al., *Efficiency Rating Systems: Their History, Organization, and Functioning* (Library of Congress, Legislative Reference Service, State Law Section, 1947), https://catalog.hathitrust.org/Record/102681008. 14.
[22] United States Bureau of Efficiency, *Report of the United States Bureau of Efficiency for the Period from March 25, 1913 to October 31, 1916*, H. Doc 1793, 64th Congress (U.S. House of Representatives, 1916), https://www.govinfo.gov/content/pkg/SERIALSET-07240_00_00-021-1793-0000/pdf/SERIALSET-07240_00_00-021-1793-0000.pdf.
[23] Classification Act of 1923, Pub. L. No. 67–516, 42 STAT 1488 (1923). § 9
[24] An Act to Amend the Classification Act of 1923, as Amended., Pub. L. No. 77–200, 55 Stat 613 (1941).
[25] Performance Rating Act of 1950, Pub. L. No. 81–873, 65 Stat 1098 (1950).

Forced ranking sounds like a good idea but the evidence is mixed at best

OPM's proposed rule makes an extended case for the benefits of a standardized distribution of performance ratings, citing both academic literature and private sector adoption. However, OPM's argument ignores both some of the drawbacks in the evidence it cites as well as the long-term move away from forced rating by even employers who had previously championed it.

To begin with the academic evidence: the Berger, Harbring, and Sliwka [26] study OPM cites does show that forced distribution can have a positive effect on productivity, but OPM elides the fact that the study is based on a laboratory experiment rather than an experiment conducted in a real employer. In this study, participants are rated on their ability "to repeatedly count the number '7' in blocks of randomly generated numbers" and the "supervisors" are randomly assigned participants. The results of the study may or may not, then, generalize to an organization as large and complicated as the federal government.

Furthermore, while the study does conclude that some performance gains are possible, it also found that:

> ...it may be problematic to set up a forced distribution when employees have experienced a more "liberal" system of performance evaluations before. Most importantly, in our within-subjects design we find that the introduction of a forced distribution leads to a short-term performance increase **which is followed by a rather sharp drop in performance.** Apparently, while the participants initially understand that they need to work harder under a forced distribution they are soon demotivated as they cannot attain the good grades and high bonuses they have earned before. [27]

Given OPM's focus on prior leniency in the system, it seems that this characterization would also apply to the federal government. This should give OPM pause in relying on this study to justify forced ranking.

To be sure, the Wijayanti et al. [28] literature review OPM references does correctly identify that there are positive benefits associated with forced ranking in parts of the literature. But OPM's presentation of this evidence is notably one-sided, and a fuller accounting of the evidence base paints a more complicated picture. For example, the authors of this literature review also

[26] Johannes Berger, Christine Harbring, and Dirk Sliwka. "Performance Appraisals and the Impact of Forced Distribution: An Experimental Investigation." IZA Discussion Paper, No. 5020 (2010).
[27] Id. (emphasis added).
[28] Aniek Wijayanti et al., "What Do We Know about the Forced Distribution System: A Systematic Literature Review and Opportunities for Future Research," *Management Review Quarterly* 75, no. 1 (2025): 747–88, https://doi.org/10.1007/s11301-023-00396-8.

conclude that "[t]he risks associated with [forced distribution] tend to outweigh the benefits when applied to jobs involving task interdependence"[29] which surely describes many jobs within the highly complex federal environment. In our view, the best way to understand the evidence base is that findings are genuinely mixed and do not point at one clear "best" approach to performance rating.

The reasons are instructive. The literature identifies not only benefits but also significant risks: underrecognition and talent loss,[30] decreased knowledge sharing,[31] organizational citizenship,[32] and so forth. These are not trivial concerns in any organizational context, and they are particularly acute in the federal government, where the work is often collaborative rather than individually competitive, where teams must function across agency boundaries and presidential transitions, and where the consequences of reduced cooperation can directly affect public safety and welfare.

Further, we'd also note that one of the most salient facts about forced ranking in the private sector is this: the companies most famously associated with the practice have largely moved away from it. Microsoft abandoned stack ranking in 2013 after years of internal criticism[33] that it encouraged destructive internal competition and discouraged collaboration.[34] General Electric—the company where Jack Welch championed the “vitality curve” that OPM’s proposed rule discusses at length—moved away from rigid forced ranking after Welch left.[35] The broader trend in the private sector over the past decade has been away from forced

[29] Id. 782

[30] Brittany M. Bond, “Cut to the Curve: Underrecognition and Talent Loss from Forced Ranking in a Multinational Firm,” *Management Science* (August 2025), https://doi.org/10.1287/mnsc.2023.01204.

[31] Linda Loberg et al., “Forced Distribution Rating Systems and Team Collaboration,” *Journal of Economic Behavior & Organization* 188 (August 2021): 18–35, https://doi.org/10.1016/j.jebo.2021.04.035.

[32] Sue H. Moon et al., “Precarious Curve Ahead: The Effects of Forced Distribution Rating Systems on Job Performance,” *Human Resource Management Review* 26, no. 2 (2016): 166–79, https://doi.org/10.1016/j.hrmr.2015.12.002.

[33] “Microsoft’s Downfall: Inside the Executive E-Mails and Cannibalistic Culture That Felled a Tech Giant,” *Vanity Fair*, July 12, 2012, https://web.archive.org/web/20141222011349/https://www.vanityfair.com/online/daily/2012/07/microsoft-downfall-emails-steve-ballmer.print.

[34] Tom Warren, “Microsoft Axes Its Controversial Employee-Ranking System,” *The Verge*, November 12, 2013, https://www.theverge.com/2013/11/12/5094864/microsoft-kills-stack-ranking-internal-structure.

[35] Max Nisen, “Why GE Had to Kill Its Annual Performance Reviews after More than Three Decades,” Economic Indicators, *Quartz*, July 20, 2022, https://qz.com/428813/ge-performance-review-strategy-shift.

distribution systems and toward simpler, technology-enabled, and less rigid rating schemes.[36] OPM's proposed rule acknowledges that one source estimates 30 percent of Fortune 500 companies use some form of standardized distribution, but the trajectory of adoption isn't straightforwardly upward.[37]

Moreover, the federal government's own experience with forced-ranking-adjacent systems has not been encouraging. Early attempts at setting extremely tight rules for high ratings in the 1950s, for example, resulted in a situation where the rules were *so* strict that performance was artificially clustered at the median rather than at the top. As a Senate report from 1952 put it:

> Congress had set in the Performance Rating Act [of 1950] very high standards for the "Outstanding" rating and the Civil Service Commission has rigidly maintained the integrity of that rating, both in the letter and the spirit. The result has been, though, that more than 99 percent of all employees in the Federal civil service are now rated "Satisfactory."[38]

The committee went on to point out that this meant that performance had "virtually no weight" in the reduction in force process as a result and impeded effective personnel management. Further, on an individual level, a letter from GAO to the committee argued that "An employee may attain an 'outstanding' rating, never having attained one before, and may never attain one again. Some look upon an 'outstanding' rating solely as a reward and not as an incentive because the prospects of attainment are so remote."[39]

More recently, the government had relevant experience via the National Security Personnel System (NSPS) at the Department of Defense which implemented a pay-for-performance system with elements of relative ranking. While NSPS did not necessarily have a rigid distribution, its mechanism for allocating bonuses was intended to be cost neutral (i.e., with no raises for low-rated employees offsetting larger raises for high-rated employees) such that an implicit distribution was baked. "[M[anagers were told to rate most employees at level three

---

[36] Julie Duda et al., "Reinventing Performance Management Processes Won't Unlock Human Performance. Here's What Will.," Deloitte Insights, March 23, 2025, https://www.deloitte.com/us/en/insights/topics/talent/human-capital-trends/2025/employee-performance-management-optimization-effective-strategy.html.

[37] Theresa Agovino, *Modernizing the Employee Performance Review* (Society for Human Resources Management (SHRM), 2023), https://www.shrm.org/topics-tools/news/all-things-work/reimagining-performance-review.

[38] *Reduction-in-Force System in the Federal Government*, S. Rept. 2102 (U.S. Senate, Committee on Post Office and Civil Service, 1952).

[39] Id.

("Valued Performer")"[40] for example, to help maintain the spread. NSPS was widely criticized for inconsistent application, perceptions of favoritism, and racial disparities in ratings before Congress terminated it.[41] OPM's proposed rule does not discuss the NSPS experience, which is a significant omission given that it represents the most recent large-scale attempt to implement performance-based differentiation in the federal government.

We also note that the private sector companies that *have* maintained some form of forced ranking (like the tech industry)[42] tend to share characteristics that the federal government lacks: significant financial incentive power (including equity compensation and large performance bonuses), substantial investment in calibration infrastructure, robust career development programs that provide genuine paths for employees rated in the middle tiers, and the organizational agility to restructure teams quickly when the system produces adverse consequences. The federal government has none of these features at comparable scale. A forced ranking system borrowed from Amazon or Meta but implemented within the constraints of the General Schedule, the federal benefits structure, and the current state of supervisory training is not the same system—it is a different system operating under different conditions, and there is no reason to expect it will produce the same results.

None of this means that forced ranking is categorically wrong for the federal government. It means that the evidence is genuinely mixed, that the implementation context matters enormously, and that confidence in the policy's likely effects should be proportional to the quality of the evidence—which, at present, does not support the certainty with which OPM's proposed rule proceeds. OPM should be more open in its final rule about the limits of the evidence base and about how much we have yet to learn about the application of new strategies to the federal government.

A better way: rigorously evaluate demonstration projects

It is not surprising that the proposal to impose forced ranking across the federal workforce has generated strong reactions on all sides. Proponents see it as a long-overdue corrective to a system that has tolerated mediocrity for decades. Opponents see it as a blunt instrument that

[40] Anthony R. Crain, *The Brief, Eventful History of the National Security Personnel System*, Occasional Paper 1 (Office of the Secretary of Defense Historical Office, 2017), https://history.defense.gov/Portals/70/Documents/occasional_papers/NSPSBOOK-FEB9WEB.PDF?ver=2017-04-17-134923-437.

[41] Jack Moore, "The Lessons and Legacy of NSPS," *Federal News Network*, January 18, 2023, https://federalnewsnetwork.com/all-news/2012/01/the-lessons-and-legacy-of-nsps/.

[42] Sarah Jackson, "Tech Companies like Amazon and Meta Are Tightening Their Performance Reviews. Here's What That Could Signal, According to Experts," CNBC, February 10, 2026, https://www.cnbc.com/2026/02/10/tech-companies-like-amazon-and-meta-are-tightening-their-performance-reviews-heres-what-that-could-signal-according-to-experts.html.

will demoralize the workforce and be gamed by supervisors. Both sides are arguing from their priors, because neither side has strong evidence about what will actually happen when this specific policy is implemented in this specific context. That is the core problem, and it is a solvable one.

The federal government already possesses statutory authority, under 5 U.S.C. § 4703, to conduct demonstration projects that test alternative personnel management approaches with appropriate controls and independent evaluation. [43] This authority exists precisely for situations like this one: where a proposed reform is plausible, but where the implementation risks are significant enough to warrant pilots and study. This authority is intended to let OPM test new ideas while also avoiding a situation where the cost of a failed government-wide rollout is borne by millions of employees all at once.

We recommend that OPM, rather than finalizing a government-wide standardized distribution requirement, instead design and implement a rigorous demonstration project—or a set of demonstration projects across agencies with varying mission types and workforce compositions—to test whether a standardized distribution of ratings actually produces the outcomes the proposed rule assumes it will. Such a demonstration should include, at minimum:

- **Randomization.** Agencies or organizational units should be randomly assigned to treatment and control conditions to isolate the effect of the standardized distribution from other concurrent reforms. Without randomization, any observed change in ratings distributions could be attributed to the Hawthorne effect, changes in supervisory behavior driven by heightened scrutiny, or other confounding factors.
- **Independent evaluation.** The assessment of results should be conducted by an independent entity—GAO, a federally funded research center, or a competitively selected research organization—not by OPM itself. OPM has an institutional interest in the success of its own policy, and allowing it to serve as both architect and evaluator of the reform would undermine the credibility of the findings.
- **Measurement of the outcomes that matter.** The relevant question is not whether a standardized distribution produces a different distribution of ratings—by definition, it will. The relevant questions are whether it improves actual organizational performance, whether it increases employee perceptions that performance is recognized meaningfully, whether it reduces or exacerbates racial and gender disparities in ratings, whether it affects retention of high performers, and whether supervisors are able to implement it in a way that employees perceive as fair. These outcomes should be measured with validated instruments over a sufficient time period to capture both short-term disruption and longer-term equilibrium effects.

[43] 5 U.S.C. 4703

- **Sufficient duration and scale.** The demonstration should run for multiple performance cycles and should be large enough to detect meaningful effect sizes across different agency types. A demonstration limited to a single agency or a single performance cycle would not generate the evidence base needed to justify—or reject—government-wide implementation.

This approach is the strongest possible foundation for a durable reform. If forced ranking works in the federal context, a rigorous evaluation will demonstrate that — and the case for government-wide adoption will rest on something far sturdier than the theoretical arguments available today. If it does not work, or if it works in some agency contexts but not others, the evaluation will surface that too. Either way, the government will have learned something real rather than imposing a policy whose effects can only be guessed at.

We recognize that this recommendation may frustrate those who believe the current system is so broken that any change is worth trying immediately. We are sympathetic to that impulse — the data on rating inflation is genuinely damning, and the status quo serves no one well. But urgency is not a substitute for rigor, and there is much to learn before we push this update to the entire workforce. This time, OPM should have the courage to ask these hard questions that its predecessors have sought to avoid.