

Written Statement for Congressional Testimony Before The House
Subcommittee on the Federal Workforce, Postal Service, and the District of
Columbia, and The Senate Subcommittee on Oversight of Government
Management, the Federal Workforce, and the District of Columbia

Barry Jay Seltser
May 22, 2007

I would like to begin by clarifying that I am submitting this statement at the request of the Committees. Because of my opposition to how some key aspects of the new performance system were being implemented, I resigned from GAO in June 2006. Since then, I have not sought out opportunities to speak about why I left. I have agreed to testify in the hope that some of my perceptions and experiences can help strengthen this important organization. Although I have not been at GAO for the past year, I believe my experiences and concerns continue to be relevant to the current performance management system.

I had a wonderful 16 year career at GAO, and greatly enjoyed the opportunities I received. I was treated very well, was promoted several times (eventually to the senior executive service), and was not negatively affected financially by any of the recent personnel changes. Also, I want to emphasize that I am not calling into question the motives or good intentions of anyone at GAO. I had, and continue to have, the highest respect for the outstanding staff and senior managers throughout the organization.

I am speaking only for myself, although many of my views are shared by some of GAO's strongest and most successful staff. My perceptions and conclusions are based both upon my own experience, and upon discussions with hundreds of GAO employees and dozens of GAO managers over the last few years. Because of my role in a specialist Team that provided advice and support across GAO, because of my level in the organization, and because I led the Task Team for Specialists that considered and made recommendations for the 2005-2006 pay

category decisions, I had the opportunity to talk with people across the organization, and most of them expressed concerns about the issues I am addressing here. I did not conduct a scientific survey of staff opinions, but I can attest, on the basis of my knowledge and experience, that these concerns are widely held and are not restricted to lower-performing or disgruntled employees, nor are they limited to staff in one or two Teams or at one or two levels of the organization.

In this statement, I will identify several closely related and problematic features of the new performance management system. The goals of pay for performance are reasonable and appropriate; the problems have emerged from specific decisions taken in the design and implementation of this particular system at GAO, leading in many cases to a lack of transparency and trust. In order to understand many of the problems that have emerged more recently with the decisions to split the largest pay band, it is necessary to take a broader view of the overall performance management system.

First, the new system is based on ratings on a group of “competencies” (such as achieving results, thinking critically, and communicating orally and in writing). Using input from staff and other management personnel, supervisors provide these ratings, which are then reviewed by unit SES managers. In theory, employees are rated on each of the competencies, and the resulting sum of the individual ratings determines their relative ranking within their unit. This relative ranking within the unit is used to allocate most of the financial and promotional rewards.

In practice, however, the actual rating scores usually have to be managed by the unit SES members to assure that the stronger staff receive a higher total rating score. Most GAO managers with whom I spoke acknowledge that the competencies are not all equally important for all employees; some staff members may be expected and required to perform particularly well on written

communication, for example, while other employees may be focused on dealing with congressional staff or oral communication. In addition, I believe most GAO managers value certain competencies such as “achieving results” and “written communication” as particularly important competencies for most of their staff members, and employees who tend to excel in those areas would be more likely to be viewed as more outstanding overall. (In some specialist units, on the other hand, competencies such as “critical thinking” and “collaborating with others” may be seen as more essential to the role.)

However, the system weights the competencies equally, even though, as I will discuss below, there appears to be no clear basis for doing so. Because of this problem of equal weighting, the subjectively correct rank ordering can be achieved only by a separate process assuring that the best employees are at the top of the distribution. This usually involves raising or lowering particular individual rating scores to assure that the summed total for the employee places him or her at the appropriate place in the overall rating distribution within the Team. On the basis of my conversations with dozens of supervisors, and with senior managers over a period of several years, I am convinced that such “managing” of the ratings was a common and essential aspect of the determination of the final rating distribution within each unit.

In my opinion, given the design of the system, there is nothing wrong with this approach, because some process of “managing ratings” is inevitable and justifiable in any performance management system. In many cases, an employee with outstanding skills and performance in one or two competency areas would be seen by the manager as making a more significant contribution overall to the Team and to GAO, and therefore more worthy of additional rewards for that rating year, than an employee with moderately strong skills in several areas. But simply adding up equally weighted scores across the competencies may not yield the appropriate rank ordering. The system assumes that the definition of “strong performance” is completely and universally identified with having a higher sum of

individual scores, and that all GAO employees at the same level are expected to perform tasks represented equally across all the competencies. I believe these are not reasonable assumptions, given the diversity of roles and the types of accomplishments that need to be considered in rewarding staff.

As a result, fairness and accuracy frequently require additional management flexibility to assure that the strongest overall performer will end up at the top of the distribution. The current system forces the rating distribution to be used as the sole determinant of most financial rewards, and usually of promotion potential as well. Instead of allowing managers to look at the rating distribution and then use their judgment to make the final rank ordering (which is how the prior performance management system operated), the current system requires managers to build the subjective judgment into the assignment of the rating scores to assure the appropriate outcome. Such subjective judgment is always an essential aspect of any such performance management process; the only question is where and when it occurs, and how visible it is to staff. My concern is that public statements about how ratings are determined have not been consistent with actual practice.

Furthermore, this process of frequently managing the rating scores undermines the usefulness of the system for purposes of providing competency-specific feedback to staff. If an employee's ratings are determined in part by a process that does not simply allow the scores to be assigned strictly according to perceived performance separately on each competency, supervisors are less able to use the ratings to provide honest feedback on staff performance. This feature of the system, I believe, both distorts the ostensible and stated meaning of the ratings, and fosters distrust among staff about the basis for the ratings themselves.

Second, to the best of my knowledge, the processes of validating the rating categories, and of determining the reliability of applying rating scores, were

insufficient. I am aware of only one formal validation effort before the system was implemented. In developing the specific standards for each competency, GAO staff were provided a long list of work activities and asked whether they performed them in their current roles and levels. The results of this exercise were used to assign standards to each level of the organization. But I am not aware of any such process to determine whether the competencies, or the work standards that defined them, should be weighted equally. I reviewed the contractor's descriptions of the validation process, and no mention was made of assuring equal weighting either of the competencies or of the associated work standards. (I once raised this question in a meeting of senior managers, and the response was that "there is no evidence that the competencies should not be equally weighted". Needless to say, GAO staff would not accept such a response to questions about a policy decision from an agency they were auditing.)

I also am unaware of any efforts to seriously assess the reliability of the rating process. Standard social science practice would require a systematic examination of "inter-rater reliability" to assure that different supervisors would provide similar rating scores when confronted with the same situation. I frequently asked whether a formal and thorough testing process had occurred before the implementation of the system, and no examples were ever provided. However, after the implementation of the system, I sat through two informal brief exercises (with supervisors and with senior managers) asking us to rate staff based on written scenarios; in both of these cases, there were large differences in how the participants rated the staff member. Such results did not give me confidence in the reliability of the ratings, nor did they suggest that such testing had been systematically performed before the system was implemented.

Achieving reliability on performance ratings is a very challenging task for any organization, and some variation is inevitable due to the subjective nature of the rating process. However, because insufficient attention was paid to this issue, and because so much now rests on the rating scores, I believe the claims made

for the objectivity and accuracy of the ratings are not supported by the evidence. This provides another basis for skepticism about whether the current system accurately and fairly assigns rewards on the basis of actual performance.

Third, although it may not be apparent to most staff, there is pressure from the top of the organization to have roughly comparable scores across the different units, and to keep the ratings consistent from year to year. Small differences are allowed but unit managers work hard to assure that their staff receive ratings that are within an acceptable range. Senior managers whose Team rating scores are higher (both overall and on particular competencies) have received additional scrutiny, and have frequently passed those concerns on to the raters in their units. In addition, pressure is exerted to avoid “rating inflation” from one year to the next, which creates additional distortion to the accuracy of the ratings. Because there is relatively low turnover at GAO, and because the written standards are not revised upward each year, I would expect ratings to increase over time, as staff became more proficient in meeting the standards; but the strong pressure to avoid inflation results in individual ratings that are often lower than deserved. Supervisors and managers throughout GAO told me about examples of being pressured to lower ratings, not because the particular rating was not correct but because the unit’s overall scores needed to be lower. This procedure undermines useful feedback and trust, since many staff receive ratings that do not reflect their actual level of performance on specific competencies.

Part of the problem is the assumption that all units are roughly equal in terms of the overall distribution of performance of their staff. Staff in “stronger” units are significantly disadvantaged, since they might be higher in the distribution ranking if they were judged against colleagues from other units. These discrepancies in the application of the ratings would be less of a problem if the relative standing within a unit was not used as the major (and often the sole) basis for determining

raises, and as a major criterion to determine placement in the upper level of the restructured "Band II" pay ranges.

This leads to consideration of what I believe is the most serious problem with the implementation of GAO's new performance management system. Much greater emphasis is now placed on a particularly inflexible approach to relative (as opposed to absolute) performance, where the reference group for the relative performance is the individual's Team (e.g., Homeland Security and Justice, or Physical Infrastructure). The major institutional rewards (including raises and promotions, as well as many opportunities for more challenging work) are increasingly based on the relative score one receives on the annual assessments. One further indication of this change is the significant decrease several years ago in the available funds for ad hoc incentive awards, when much of the money was put instead into the annual evaluation process. The more ad hoc awards could be used by managers to provide incentives to staff who may have ended up in the lower part of their rating distributions but who still made very significant contributions; decreasing the availability of these funds sent a further message that staff whose relative performance within their work unit was not high were not highly valued by the organization.

My major concern is that these changes have created a much more competitive environment at GAO, where an employee's financial opportunities and organizational status are being competed in a "zero-sum" atmosphere where only a certain percentage of staff will ever be treated as strong performers. A manager at GAO is now much less able to reward a staff member who has worked hard and made significant accomplishments on particular projects, unless the staff member is placed in the top part of the rating distribution within the Team. The basis for this approach appears to be a belief that a relatively small group of employees in each unit consistently perform much better than their peers, and that they should receive the major rewards of the performance management system.

But this belief may not be accurate for all Teams, and indeed is unlikely to be equally true in diverse units with different types of priorities and projects. In some Teams, the contributions of staff may be much more homogeneous, in which case a true “pay for performance” system should be rewarding people more equally. It is, I believe, the task and responsibility of senior managers within the Team to make these determinations, and they should be given the flexibility to provide incentives and rewards according to the actual distribution of performance within the Team. The current system does not allow such flexibility. For example, when the decision was made to place staff in either the 2A or the 2B pay categories, the midpoint of the rating distribution was used in all Teams as one of the key factors, a decision that clearly disadvantaged staff in stronger Teams and that failed to recognize different levels of performance and skills among people who might have ended up in the bottom half of the rating distribution in many Teams.

This approach has demoralized many excellent GAO employees who are making very significant contributions, but who may not be as strong as their peers in their particular unit in overall performance when senior managers are required to rank order their staff. As a Director of a specialist unit, for example, I created a rating distribution based on relatively small distinctions in staff performance; but those in the bottom half of the distribution were then treated in significantly different ways and were sent messages that they were not strong performers, in spite of their often enormous contributions to the organization. I do not believe this is an effective or helpful way to implement the concept of “pay for performance”, in part because it over-emphasizes relatively small differences in staff performance and because it sets staff against one another in terms of competing for rewards within their own work unit. GAO’s work depends upon teamwork, collaboration, and mutual support from one’s colleagues; any incentive system that threatens to undermine teamwork does not serve the interests of either Congress or the American people.

Finally, I was very concerned by some other aspects of how the decisions were made to assign staff to the newly created pay categories in 2006. The decisions were based primarily (although not exclusively for most analyst staff) on ratings given during the prior three years; however, when these ratings had been assigned, supervisors and managers had no idea that they would be eventually used in this manner. For most specialists at GAO, their ratings were the only relevant factor determining their pay category in this decision, and staff who were fully qualified to perform functions and roles defined in the criteria for the higher category were placed in the lower one solely because they had been in the bottom half of the rating distribution for their work units. This result was not consistent with the original statements that the placements would be based primarily on roles and responsibilities, and that no fixed quota would be used to determine the percentage of staff in the higher pay category.

This outcome placed managers in a very difficult position of justifying using ratings for a purpose for which they were never intended. An individual's overall relative ranking on the competencies was not, in my view, a predictor of the ability required to perform at the higher pay level. If the decision to assign most of the affected staff to the lower pay category was a purely financial one (because GAO could not afford to pay more than half of its staff at the cap of the higher pay category), this rationale should have been clarified at the outset of the process. Instead, various reasons were provided, the major one being that roles and responsibilities should be more highly correlated with pay. But no convincing argument was made that one's relative rating scores should be used as a proxy for such decisions, particularly if the distribution of skills and accomplishments varied across work units and across types of staff.

In addition, the process for making the final placements was taken out of the hands of the unit managers who were closest to staff performance, and who would be able to understand and assess their skill levels and experiences most

effectively. In the interests of creating a process that would be seen as sufficiently consistent across the organization, the unit managers' views were frequently overridden. The Executive Committee has the ultimate right to make such decisions, of course, but I believe it did a serious disservice to the broader management team and to the expectations of staff who felt that their unit managers knew more about their actual performance and their abilities to meet the other criteria of the higher pay category.

In conclusion, I believe that many of these problems could have been avoided by different decisions that would have preserved the positive goals of implementing and improving GAO's earlier pay for performance model. For example, more openness about the actual process of assigning rating scores, including recognizing the subjective nature of how managers assign ratings and the need to weight some competencies more highly than others, would have created more trust in the underlying scores, and less skepticism about how management was defending and using the ratings.

In addition, more flexibility in how distinctions were drawn in different work units would have provided managers with the ability to apply the ranking system in a way that truly matched the performance distinctions of their employees, and would have prevented staff from being seriously disadvantaged by being part of a stronger work unit. One alternative would have been to provide unit managers with a per capita budget for raises and other rewards, and to allow them, with some guidelines, to apportion the rewards according to the actual performance of their staff. That is, after all, the underlying purpose of a pay for performance system, which should focus on rewarding actual performance levels rather than on establishing fixed ranking levels to determine reward differentials. In a unit where skills and accomplishments are more homogeneous, for example, the rewards should be spread more evenly as well. Using a fixed cutoff level or expecting similar rating distributions across all work units undermines the ability of managers to create a reward structure that reflects the actual significant

differences in performance of their employees. GAO's senior managers are paid for their judgment and professionalism; a process that minimizes their flexibility and ability to tailor rewards to performance is likely to undermine both their self-respect and the way they (and the performance management system) are perceived by their staff.

It is possible that GAO's manner of designing and implementing this performance management system was the only way to avoid legal problems or other unanticipated consequences. I do not believe that this is the case, but if it is, then I would hope Congress, in its oversight function, would be fully aware of the effects and implications of adopting such a system in other governmental agencies. The problems are serious; if they cannot be avoided, they should at least be examined, acknowledged, and minimized wherever possible.

I believe that GAO's current performance management system includes features that are distorting the underlying goals of a pay for performance system, and that the actual implementation of this system threatens to undermine the teamwork and collaboration upon which GAO's work depends. During my final years at GAO, I observed a serious deterioration in morale and trust as a result of these changes, including among some of the very best employees at GAO. As a senior manager, I was forced to send negative messages to some excellent employees who fell into the bottom half of the rating distribution in a very strong unit. At the same time, many others grew increasingly uncomfortable with a system that, while providing more financial rewards to them in some cases, clearly harmed their colleagues, and seemed to be creating a more competitive and unfair work environment for all employees. In my view, the gains from this approach were not worth the harm that was done to many outstanding employees and to the culture of the organization.

The current Comptroller General and his Executive Committee have made many positive changes at GAO, and the underlying goal of a system that rewards high

performance is one that should be fostered. But I believe that Congress needs to give careful consideration to the specifics of how this system is being implemented, and the effects this may have on the perception of fairness and on the morale and teamwork upon which government work depends. GAO's experiments in the human resources area are frequently cited as models for the rest of government, so it is very important that Congress understands what GAO has done and how to avoid some of its problems. I hope that GAO, and the rest of the federal government, can move toward a version of performance-based rewards that take into account both the benefits and risks of implementing this approach in the all-important arena of public service.