

Appraising Performance Appraisal Systems in the Federal Government:
A Literature Review, Preliminary Findings, and Prospects for Future Research

Dr. Ellen V. Rubin
Department of Public Administration and Policy
Rockefeller College of Public Affairs and Policy
State University of New York, University at Albany
314B Milne Hall
135 Washington Ave
Albany, NY 12222
518-442-5261
erubin@uamail.albany.edu

Presented at the Public Management Research Conference
June 2011
Syracuse University

The author welcomes your comments. Cite only with permission of author.

Appraising Performance Appraisal Systems in the Federal Government:
A Literature Review, Preliminary Findings, and Prospects for Future Research

Abstract

Performance appraisal systems are a key tool for holding civil servants accountable and should be considered as a key variable of the larger performance management equation. In recognition of this, the Office of Personnel Management (OPM) and the Government Accountability Office (GAO) each promulgated criteria for effective appraisal systems. Analyzing the OPM data through the lens of the GAO criteria reveals that less than half of the audited appraisal systems meet GAO standards. Specifically, appraisal systems are most effective in involving employees and providing feedback on performance. Conversely, systems are weakest in differentiating between levels of performance and in providing consequences for performance ratings. These results suggest analyzing performance appraisal system structures, beyond employee survey data, provides scholars with rich research opportunities.

Appraising Performance Appraisal Systems in the Federal Government:
A Literature Review, Preliminary Findings, and Prospects for Future Research

Performance appraisals are a key tool for holding individual public servants accountable. Whether we choose to evaluate the “steering” or “rowing” of individual bureaucrats, we do so with appraisal systems, which are usually required by law. Despite this, the literatures on accountability and performance management often ignore the potential of appraisals.

Federal agencies have expended significant time and effort updating performance appraisal systems over the last decade. Under the President’s Management Agenda, agencies were encouraged by both the Office of Personnel Management (OPM) and the Government Accountability Office (GAO) to reform and reinvigorate performance appraisal systems. The GAO and OPM drumbeats for reform were consistent with the themes of the performance management movement and New Public Management more broadly. In particular, GAO recommends that performance appraisal systems should “make meaningful distinctions in performance.... align individual performance expectations with organizational goals.... and identify the competencies that individuals need to accomplish organizational results” (GAO 2003, 4). Appraisal systems exhibiting these characteristics can be expected to enable the creation of a results-oriented organizational culture, according to GAO.

Much of what we know about federal appraisal systems, however, is based on multiple strobos of employee survey data. These surveys, conducted by OPM and the Merit Systems Protection Board, are valuable markers of employee perceptions of the systems. For example, in the most recent iteration of the OPM survey, 68 percent of federal employees believe performance appraisals fairly reflect individual performance (OPM 2010). What we lack is systematic knowledge about the actual rules and structures of these systems. This research aims

to close this gap by examining data collected by OPM on the structure of agency appraisal systems. By combining both rules and perceptions, scholars obtain a more complete view of performance appraisals, enabling more robust empirical analysis.

The central aim of this paper is to analyze the degree to which federal agency appraisal systems meet the GAO criteria for effectiveness. To answer this question, data from OPM audits of agency appraisal systems are organized and assessed to compare current systems against the GAO standards. Analyzing the OPM data through the lens of the GAO criteria reveals that less than half of the audited appraisal systems meet GAO standards. Specifically, appraisal systems are most effective in involving employees and providing feedback on performance. Conversely, systems are weakest in differentiating between level of performance and in providing consequences for performance ratings. Furthermore, half of the agencies that participated in multiple OPM audits made significant improvements to their appraisal systems.

To review these results more fully, this manuscript begins with a brief review of the history of performance appraisal systems, a review of the public administration literature on performance appraisals, and a brief consideration of the parallels between performance appraisals, accountability, and the performance management literatures. The development of the GAO and OPM standards for performance appraisal systems is then chronicled, followed by a description of the data set provided by OPM. Once the findings have been detailed, the paper closes with a discussion and proposals for future research.

A Brief History of Public Sector Performance Appraisal

Bureaucracies have been evaluating employee performance for thousands of years. Chinese civil servants and military officers underwent mental, moral, and physical fitness evaluations as far back as 200 BCE (Weise and Buckley 1998). Even at this early date,

evaluators noted the difficulty of conducting objective evaluations. In the Middle Ages, European Guilds used evaluations for certifying craftsmen as Masters, and early universities used exams to evaluate students of divinity and the liberal arts (DeVries et al 1981). The earliest record of performance appraisal usage in the industrial ages dates to the early 1800's. At the end of each working day, Scottish cotton mill workers were rated on performance, and the ratings were posted above individual workstations for all to see (DeVries et al 1981).

In the U.S., early appraisal activities were conducted by governments and the military. Wars and the concurrent expansion of the civil service frequently motivated the promulgation of performance appraisal systems. As the government expanded spending, elected officials and the public became more interested in government efficiency, while managers within the bureaucracy sought ways to document performance for the purposes of promotions and downsizing. For example, after the War of 1812, the Army General documented evaluations of soldiers and retained the records centrally (Weise and Buckley 1998). This effort later expanded to conduct “efficiency ratings: of the civilian federal workforce:

- 1887 President Harrison directs agencies to document the efficiency of employees;
- 1912 Civil Service Commission creates of the Division of Efficiency, but regulations only cover federal employees located in Washington, D.C.;
- 1923 Classification Act requires efficiency ratings for classified civil servants;
- 1940 and 1941 Ramspeck Act and amendments revise efficiency ratings requirements, and require agencies to create panels to hear appeals of efficiency ratings and rate the efficiency of employees outside of Washington, D.C.;
- 1946 employees are granted the right to appeal efficiency ratings to the Commission and rating systems must be approved by the Commission; and

- 1950 Performance Rating Act requires a three-tier rating system (Van Riper 1958).

More recent legislation influencing performance appraisal includes the Civil Rights Act of 1964 and the Civil Service Reform Act of 1978. The Civil Right Act of 1964, and related court cases and regulations, impacted appraisal in both the public and private sectors. Under the law, organizations are required to validate appraisal systems to ensure they measure criteria that are related to job performance and to ensure appraisals are not discriminatory in intent or practice (Daley 2004). Agencies can not simply assert appraisal systems are valid and non-discriminatory; they are required to document the appropriateness of systems in a manner that could withstand judicial scrutiny.

A desire to hold federal employees accountable for performance served as one motivation for the passage of the Civil Service Reform Act of 1978. Once again, agencies were required to develop appraisal systems, now to be approved by the new Office of Personnel Management (OPM). However, the new appraisals were required to be consistent with the Civil Rights Act, and provide opportunities for employee participation. Appraisals were now used to inform other personnel actions such as promotions and training. Furthermore, employees could now be removed for documented performance problems if they failed to improve during officially documented performance improvement periods. Pay-for-performance for federal managers, linking appraisal results to pay increases, was also included in the Civil Service Reform legislation. Its well-documented failures (for example see Pearce and Perry 1983) eventually led to repeal of this part of the law. Despite this failure, officials remain interested in linking performance appraisal ratings to pay decisions, based on the assumption that civil servant are extrinsically motivated and would improve efficiency and effectiveness if provided with the opportunity to be rewarded.

Performance Appraisal in Public Administration Scholarship

The literature on performance appraisals is broad, covering how to design one, how to implement one at the organizational level, how managers are to carry them out at the individual level, psychometric properties of different schemes, and problems with appraisal and how to minimize them. A review of this entire literature across multiple disciplines is beyond the scope of this paper. However, it is important to understand how performance appraisal has been covered in our field.

The literature on performance appraisals at the national level can be divided into four categories. First, a series of papers focus on appraisal methods and implementation in a specific agency. These papers either describe a new appraisal system (Brumback and McFee 1982, Brumback 1993, and McNish 1986) or analyze survey data describing employees' perceptions of their agency's appraisal system (Kim 1992, deLeon and Ewen 1997). Second, articles describe systems used at the national level in governments outside the U.S. (Bourgault, Dion, and Lemay 1993; Mikkelsen, Ogaard, and Lovrich 1997). Third, a series of articles discusses the implementation of the Civil Service Reform Act of 1978 (CSRA), which required appraisal systems governmentwide. A number of these pieces summarized the results of surveys conducted before and after implementation (Nigro 1981; Pearce and Perry 1983; Colby and Ingraham 1982). Fourth, scholars employing various iterations of the Merit Principle Survey assess federal employee perceptions of performance appraisals. Perceptions of performance appraisals are used as both dependent variables (Oh and Lewis 2009; Daley 2007; Yang and Kassekert 2010) and independent variables (Brewer and Selden 2000; Brewer 2005; Daley 2008).

Examination of appraisals at the state and local level is much less diverse. Only three studies focus on appraisal use at the state level, each describing the implementation of new appraisal systems (Lovrich et al 1981; Daley 1985; Moussavi and Ashbaugh 1995). Similar to the studies at the federal level, one local government study describes the implementation of a new system (Gabris and Ihrke 2000) while other studies sample multiple local governments to understand the variety of appraisal methods being used (Ammons and Rodruques 1986; England and Parle 1987; Daley 1991).

Studies of performance appraisal systems include a surprising number of studies using quasi-experimental methods. In addition to the CSRA surveys mentioned above, two additional studies use pre/post testing to evaluate the performance of a new appraisal systems in a unnamed federal agency and a unnamed local government (deLeon and Ewen 1997; Gabris and Ihrke 2000). A third study uses a pre/post/post evaluation with treatment and control agencies to assess the effects of an unnamed state's new appraisal system which included more employee participation and goal setting (Moussavi and Ashbaugh 1995). Yet another study uses a pre/post/post evaluation method with matched pairs of agencies to assess a participatory form of appraisal in Washington state (Lovrich et al 1981).

This initial review of performance appraisal research in the top public administration journals is interesting for what it lacks. Specifically, a substantive linkage to accountability and organization-level performance management systems is absent. Furthermore, there is no research describing the variation in performance appraisal systems in either the federal government or state governments, and local level variation has gone unexplored since the late 1980s. Once scholars understand the variation, we can then link these data to other perceptual and organizational outcomes. These comparisons would allow scholars to assess the effects of

various schemes. Understanding this variety also enables scholars to consider the likely effectiveness of other policies which depend on accurate and valid performance appraisals, especially pay-for-performance schemes.

Linking Performance Appraisal, Accountability, and Performance Management

More than four decades ago, Mosher lamented the accountability challenges posed by civil servants isolated from direct contact with the public. At the individual level, performance appraisals are used as *the* accountability mechanism. Scholarship on accountability rarely addresses the issue of individual performance appraisal, although it frequently discusses the “answerability” of individual civil servants.

Individual performance appraisal schemes developed from the Weberian and Taylorite control perspective; work could be fully dissected into discrete, objective, and observable steps. This foundation is consistent with Finer’s (1941) control orientation and the “make the managers manage” ideology (Behn 2001). Standards-based appraisal systems provide managers with detailed descriptions for each criterion at multiple levels of performance. Appraisals develop into highly bureaucratic systems with detailed rules, timelines, and levels of review.

Appraisals may also rely on some degree of professional discretion to define and measure performance (Friedrich 1940; Romzek and Dubnick 1987). For example, in many government organizations, employees participate in designing the performance criteria, or verify that the criteria are appropriate to their work. In some systems, employees work directly with supervisors to develop performance criteria or draft individual development plans for the year. Both rely on the expert judgment of employees and managers to identify appropriate criteria based on professional and organizational standards.

Performance appraisal systems share many commonalities with broader organizational performance management efforts.¹ Despite New Public Management rhetoric about holding managers accountable for outcomes, these prognostications typically fail to consider the role of individual performance appraisal in the larger effort to evaluate organization success. This is a disappointing gap and a missed opportunity for a number of reasons. Specifically, individual performance appraisal and organizational performance management aims to achieve similar ends via parallel means and suffer from similar problems.

Both organizational performance management systems and performance appraisal systems face significant implementation challenges internally. One major challenge faced by both performance appraisal and performance management systems is multiple, conflicting goals. On the performance management side, the competing goals of government programs, namely, equity, fair process, efficiency, and effectiveness require competing performance measure to ensure none is ignored or minimized (Radin 2006). In performance appraisal systems, a desire identify professional development opportunities may conflict with a need to document weakness and differences in individual performance for the sake of informing other personnel decisions like raises, promotions, or discipline (Daley 2004).

Other similarities between organizational performance management systems and individual performance appraisal systems abound. Both systems are criticized as being viewed as an ends in themselves instead of as a means to the end of improved performance (Murphy and Cleveland 1995; Hatry 2007). Under both systems, we assume performance can be measured

¹ For the sake of conceptual clarity in this paper, the phrase performance appraisal will refer to assessments of the performance of individual employees, while the phrase performance management will refer to evaluating performance at the programmatic or organizational level. The author acknowledges that the phrase performance management is often used to refer to more sophisticated performance appraisal systems that include feedback throughout the appraisal cycle, professional development, opportunities for self-assessment, a linkage to the accomplishment of organizational goals, etc.

objectively, and we struggle with how to gather and process performance data (Moynihan 2008). Finally, we also assume that we know and can agree on what constitutes high performance (Murphy and Cleveland 1995; Hatry 2007).

Evaluating Performance Appraisals

Performance appraisals are used to document the performance of individuals, but how should we evaluate the effectiveness of the appraisal systems themselves? If we limit ourselves to purely academic sources, at least three options emerge. Scholars interested in evaluating performance appraisal systems could focus purely on the psychometric properties of validity, reliability, accuracy of performance measures, and degree of rater error. However, these criteria are devoid of any organizational context. Alternatively, researchers could use different criteria for each organization to determine if each system is meeting its stated goals and purpose (Murphy and Cleveland 1995). While this is responsive to organizational context, it inhibits our ability to make comparisons across organizations. A third option is a program evaluation approach, tracking perceptions, implementation, and outcomes over an annual performance cycle. Such an evaluation can include consideration of the degree of organizational alignment, employee performance at the beginning and end of the cycle, and the nature of performance-driven behavior in the organization (Silverman and Muller 2009). From this brief list, it is easy to see that: “there are a wide variety of criterion measures... [such that] the choice of any specific set of measures can be criticized as arbitrary” (Murphy and Cleveland 1995, 398).

In the federal sector there are a limited number of organizations that may promulgate and apply criteria for evaluating performance appraisal systems. Organizations with formal, legally derived oversight responsibilities on this matter can be limited to GAO and OPM and the Office of Management and Budget (OMB). In the last decade, all three agencies issued publically

available criteria for evaluating performance appraisal systems. As the following discussion will describe, the criteria are, fortunately, largely consistent with one another. The increased attention on the topic can be attributed the efforts of three actors: 1) GAO's designation of strategic human capital management as a high risk area; 2) the President's Management Agenda's emphasis on human capital management generally and performance appraisal specifically, coordinated by OMB; and 3) OPM's development of the Human Capital Assessment and Accountability Framework and the Performance Appraisal Assessment Tool (PAAT).² In fact, the work by these three actors was happening simultaneously, with the frequent exchange of ideas between key staff. While some may describe the activities as institutional competition, the efforts of each agency complemented the others with the common goal of improving the management of government personnel.

In January 2001 GAO added strategic human capital management to its list of high risk areas deserving significant attention from the Congress and the President. This designation launched a body of work examining human capital planning efforts in federal agencies, with the emphasis placed on how personnel policies and personnel management impacted, positively or negatively, the ability of agencies to accomplish programmatic goals. Importantly, work supporting the high risk designation included consideration of performance appraisal systems. The themes of GAO's work on performance appraisals were informed by lessons learned from implementation of the Government Performance and Results Act. Later, GAO (2003) proposed criteria for evaluating the effectiveness of performance management systems, including 1) aligning individual performance expectations with the accomplishment of organizational goals,

² The author acknowledges many people are uncomfortable with the phrase "human capital." However, this is the language used by OPM, GAO, and the Office of Management and Budget during the last decade. Additionally, the phrase is entombed in federal law with the creation of the position of Chief Human Capital Officers in major agencies and the formation of the Chief Human Capital Officers Council.

2) providing feedback throughout the year, 3) using competencies as performance criteria, 4) linking pay to performance, 5) making distinctions between different levels of performance, and 6) involving employees in design and implementation of performance management systems (see figure 1).

[Insert figure 1 about here]

The President's Management Agenda, formally presented in August 2001, included criteria for improving personnel management systems. The agenda on human capital management emphasized the alignment of personnel policies to support the accomplishment of organizational missions broadly. In relation to performance appraisal, the agenda directed agencies to ensure they were making meaningful differentiations between high and low performers and that there were consequences for high and low performance. This expectation applied to appraisal systems spanning line employees, management, and senior executives. On a quarterly basis, agencies were rated on both their progress accomplishing these goals and on the overall status of their efforts.

OPM released the Human Capital Assessment and Accountability Framework in 2002 to help agencies meet the standards under the President's Management Agenda. Three years later, staff within OPM initiated the development of a self-evaluation tool to assess agency appraisal systems. Consistent with the goals of the President's Management Agenda, the Performance Appraisal Assessment Tool (PAAT) asks agencies to document how appraisal systems link employee performance to organizational and programmatic goal accomplishment (OPM 2006).³ In addition to the results-oriented interest, the PAAT evaluates the types of performance criteria used; the degree to which employees are involved in the design, goal setting and assessment process; and how the agency ensures the appraisal systems is implemented. OPM as an

³ Available at <http://www.opm.gov/perform/PAAT.asp>.

institution views the PAAT as an important tool for meeting its strategic objectives; the PAAT is mentioned explicitly in the 2006-2010 and 2010-2015 strategic plans.

According to agency officials, the PAAT was initially developed in response to a legislative proposal from the Bush administration initially called the Civil Service Modernization Act of 2005, and later referred to as the Working for America Act (personal communication, February 10, 2011). This law, if it had been passed and signed, aimed to abolish the General Schedule and replace it with a pay-for-performance system, among other proposals. Before agencies could implement pay-for-performance, they would have been required to have their appraisal systems certified by OPM. The PAAT was initially thought of as a potential tool for certifying agency appraisal systems. Prior to the use of this tool, performance appraisal systems were only evaluated via compliance audits and OPM lacked any understanding of the details of various approaches (personal communication, February 10, 2011).

The PAAT evaluates appraisal systems against ten equally weighted criteria: 1) alignment, 2) results focus, 3) credibility of measures, 4) distinctions in levels of performance, 5) consequences, 6) employee involvement, 7) feedback given to employees, 8) training, 9) organizational assessment and guidance, and 10) oversight and accountability (see figure 2) (OPM 2006). According to agency officials, these criteria were informed by the Human Capital Assessment and Accountability Framework, the President's Management Agenda, and the GAO human capital work. Guidance from a host of existing statutes also shaped the PAAT criteria, including the (now defunct) Defense and Homeland Security personnel reform legislation, legislation creating the (now defunct) Human Capital Performance Fund, and guidelines for reforming the appraisal system of Senior Executive Service members (personal communication,

February 10, 2011). The PAAT was pilot tested at twenty-one beta sites in late 2005 and early 2006.

[Insert figure 2 about here]

Examining the two figures makes it clear that there are differences between the GAO and OPM criteria. First, OPM's PAAT neither requires nor recommends agencies rely on competencies to assess individual performance. Instead, competencies are mentioned as one of many potential measures for employee performance, and as a tool for improving managers' implementation of performance management. GAO (2003) defines competencies as "skills and supporting behaviors that individuals are expected to exhibit to carry out their work effectively" (15) and provides examples of public sector organizations that use competencies to hold employees accountable.

Second, while the PAAT includes questions regarding the consequences of appraisal ratings, it does not advocate or assess the degree to which agencies engage in pay-for-performance. Agencies are asked to report the use of one-time cash awards, time-off awards, and quality step increases (OPM 2008). According to the scoring sheet, agencies earn similar points for effectively addressing poor performance as they do for providing various awards. Conversely, GAO (2003) explicitly argues in support of pay-for-performance systems.

It is worth acknowledging here that individuals likely disagree with GAO's criteria for effective performance management systems, particularly the two items just discussed. For example, there is much evidence, including from the federal sector, that pay-for-performance does not work and may in fact decrease motivation among public employees (for example see Pearce and Perry 1983; Oh and Lewis 2009). Likewise, scholars disagree on whether

competencies are the latest performance management fad or something more enduring (for example see Page, Hood, and Lodge 2005; Getha Taylor 2008).

Data and Methods

As noted above, the public administration literature lacks an overview of the variations in federal performance appraisal systems. Such information is worth having because of the direct link to discussions about broader performance management and accountability. The PAAT data, provided by OPM in February 2011 to the author, are used to understand the degree to which federal performance appraisal systems are consistent with the GAO criteria of aligning individual performance expectations with the accomplishment of organizational goals, providing feedback throughout the year, making distinctions between different levels of performance, and involving employees in design and implementation of performance management systems. Consequences of performance appraisal ratings will also be considered in this analysis. While GAO focuses exclusively on pay-for-performance, OPM considers a broader set of recognitions and punishments. Assessment of the use of competencies is not possible because the PAAT does not explicitly collect data on this.

To implement the PAAT, agencies apply the OPM criteria via self-assessment. An agency's score for each of the ten criteria described above are assessed based on an agency's answer to multiple supporting questions. Each of the ten criteria are supported by three to eight multi-part questions for the agencies to answer and provide supporting documentation. Each criterion is assigned a fixed number of points. The criteria are transparent to both the public and the agencies, and the criteria have remained consistent since 2005. OPM then scores agency responses to develop a measure of the effectiveness of the appraisal system. Four separate OPM experts in performance appraisal systems evaluate each PAAT submission for inter-rater reliability. Agencies are provided with detailed scores so they understand how many points were allocated to different criteria and where they fell short in the eyes of OPM. OPM designates a

system as effective when an agency receives a score of 80 percent or higher (personal communication, April 23, 2010).

Importantly, agencies are not required by any law or regulation to participate in the PAAT process. Similarly, there is no law or regulation dictating how frequently they should participate (annually, bi-annually, etc.) or that everyone participating should submit their documentation at the same time. Participation is voluntary and requires a significant amount of effort to complete the assessment. Agencies with multiple appraisal systems across multiple sub-components face even greater burdens. Those who do participate make an affirmative choice to do so, either due to peer pressure, persistent encouragement from OPM, a desire to be recognized as a leader, or for a desire to get feedback, suggestions, and/or ammunition to make improvements. As a result, some agencies have chosen to be completely evaluated multiple times, and some agencies have only had appraisal systems covering small portions of their employees reviewed once.

This essay will use the same threshold of 80 percent to designate an agency as successfully meeting the GAO criteria, or a being effective according to the GAO rubric. The data reflect information collected by OPM as of February 2011. Some agencies have submitted multiple PAATs since the initial pilot tests in late 2005; only the most recent PAAT is analyzed here.

Reliance on the PAAT information creates a number of data limitations. First, many federal agencies have not participated in the PAAT process. Agencies with significant gaps include Defense and Homeland Security, among others. The two departments include a sizeable portion of the federal population. However, the gaps are not entirely surprising considering both departments spent a significant part of the 2000's designing new appraisal systems as part of

larger personnel reform efforts, only to have Congress withdraw funding and then cancel the programs. It can be assumed that a majority of employees in both departments are currently working under the legacy appraisal systems. A second data limitation is attributed to the nature of the data provided by OPM to the author. The database on which the analysis here is based contains the scores assigned by OPM and very brief summary notes. At the present time, the researcher has not been granted access to the full documentation on which the scores are based. This limits the details that can be provided for more in-depth explanation of scores.

Determining the degree to which federal performance appraisal systems are consistent with the five GAO criteria of aligning individual performance expectations with the accomplishment of organizational goals, providing feedback throughout the year, making distinctions between different levels of performance, involving employees, and realizing consequences for performance appraisal ratings requires the use of descriptive statistics. Performance against GAO's first criterion of alignment will be assessed by combining OPM's criteria of alignment and results-focused orientation. The GAO criterion on distinguishing levels of performance will be assessed by OPM's criterion of distinctions in levels of performance plus one point added if the agency uses a rating scale other than pass/fail for individual performance elements. The other three GAO criteria map directly to OPM criteria unmodified by the author.

Findings

It is clear there is a broad distribution of the quality of performance appraisal systems in the federal government. The average agency score on the GAO criteria is 67 percent with a standard deviation of 0.229. Before examining each criterion in detail, it is appropriate to examine the degree to which agencies meet the GAO standards according to the 80 percent rule, explained above. Sixty out of 138 agencies, or 43 percent, use appraisal systems consistent with

the GAO criteria (see table 1 and figure 2). This includes fifteen agencies scoring 90 percent or higher, such as the Department of Commerce, the OMB, and the Office of the Comptroller of the Currency. Conversely, seven agency systems earned ratings below 20 percent.⁴ This finding, while somewhat disappointing, provides new insights. Specifically, existing knowledge about the status of federal appraisal systems is limited to perceptual survey data, gathered by OPM and the Merit Systems Protection Board in their regular temperature-taking of employees. The scores used here, however, include both survey results and information on the structure of the appraisal systems.⁵

[Insert table 1 and figure 3 about here]

Evidence is also available to support the argument that agencies spent significant resources revising their appraisal systems over the last decade. To make this determination, the scores of agencies submitting multiple PAATs were evaluated for changes greater than 20 points from the initial score to the most recent score. The decision rule of 20 indicating substantive change is applied here as a conservative indicator. Some change in agency scores can likely be attributed to agency learning; agencies likely learned from the initial feedback received from OPM and had the opportunity to participated in workshops and networking events that would have provided clues on how to improve the summary of the systems, regardless if changes had actually been made. Of the 60 agencies submitting multiple PAATs, 35 improved their scores by 20 points or more. Among these, seven agencies improved their score by more than 60 points,

⁴ The author acknowledges it would be useful to put the number of agencies with effective systems in context, specifically the percentage of federal employees covered by effective appraisal systems. Agencies filed PAATs during different years. As a result it is impossible to calculate the percent of employees covered by systems scoring 80 percent on the GAO criteria. Given the limitations of the data, it is only possible to calculate the percent of employees in the particular agency in the time period the PAAT was filed.

⁵Questions supporting the PAAT criteria include consideration of employee survey data.

including the Smithsonian, the Transportation Security Administration (TSA), and the National Institutes of Standards and Technology.

The details behind the changes at TSA and the Smithsonian are particularly interesting for their similarity, despite the differences in their respective workforces. Both systems now achieve or exceed the 80 percent threshold (see table 2). To accomplish this rating, TSA and the Smithsonian improved their scores on all of the criteria. For example, both agencies now document that they have appraisals that align with organizational goals and require at least one critical performance element linking to organizational performance. Four rating levels are used in both agencies, but the labels for the four levels are different. Across these four rating levels, the TSA and Smithsonian demonstrate employees are rated at all levels, i.e. ratings are not “inflated” to the top and not everyone is rated the same. Furthermore, the two agencies provided evidence that a great majority of their employees receive a mid-cycle progress review and described the system they have in place to track whether these conversations actually occurred.

[Insert table 2 about here]

Agencies are least effective in the areas of providing consequences for performance ratings and making distinctions based on performance. The number of agencies scoring at or above 80 percent is lowest for the criterion regarding the consequences for performance ratings (26 out of 138 agencies) and making distinctions in performance (18 agencies). Furthermore, the average score across all participating agencies for consequences for performance ratings is 59 percent (std. dev. = 0.222) and 55 percent for making distinctions in performance (std. dev. = 0.254). This is the lowest average score of the five GAO criteria. Just three agencies scored 100 percent in providing consequences: the Patent and Trademark Office, the Bonneville Power Authority appraisal system covering wage grade non-supervisors, and the Community Relations

Service in the Department of Justice. Only the Department of Commerce (excluding the Patent and Trademark Office and the National Institute of Standards and Technology) and the Federal Rail Road Administration earn 100 percent ratings for differentiating between performers. Importantly, this is consistent with repeated survey findings suggesting agencies do not address poor performance and that rewards and recognition are not significantly based on performance (for example see OPM 2010). As a result, scholars now have both perceptual and systems evidence that these are areas require additional attention from management.

Alternatively, agencies are most effective in the areas of providing feedback and involving employees in the design and administration of appraisal systems. Employee involvement demonstrates the largest number of agencies achieving the 80 percent threshold, with 89 out of 138 agencies earning that designation. Similarly, 81 agencies achieve the 80 percent threshold in providing systematic feedback on performance. The average governmentwide score for employee involvement is 78 percent (std. dev. = 0.220) and 71 percent for providing feedback (std. dev. = 0.306). Furthermore 37 agencies earned 100 percent ratings for employee involvement and 47 agencies earned the same for providing feedback. Admittedly, an agency saying it has rules about employee involvement and providing feedback is not the same as actually engaging in both substantively. However, the presence of these rules and systems support, reinforce, and enable meaningful participation.

The criterion of alignment represents the operationalization of the managing-for-results movement in the performance appraisal system. To borrow a phrase from GAO, it is under this criterion that federal employees are provided a “line of sight” from their daily work to the accomplishment of organizational and programmatic goals stated in strategic plans. On average, agencies earn a score of 64 percent (std. dev. = 0.337) in the area of alignment. Just less than

half of agencies achieve the 80 percent target, and only 16 earned a 100 percent rating. It is difficult to determine if this moderate performance is reasonable or troubling. Federal agencies are required by law to prepare and maintain strategic plans and report their performance at the organizational level, so agencies cannot say they lack goals with which they can link. The moderate performance on this criterion is less than what would be expected if researchers relied solely on employee perceptions which suggest that a great majority of civil servants understand how their job links to organizational performance (OPM 2010).

Discussion and Conclusion

The central aim of this paper was to analyze the degree to which federal agency appraisal systems meet the GAO criteria for effectiveness. To answer this question, data from OPM audits of agency appraisal systems were assessed to compare current systems against the GAO standards. Analyzing the OPM data through the lens of the GAO criteria reveals that less than half of the audited appraisal systems meet GAO standards for effectiveness. Specifically, appraisal systems are most effective in involving employees and providing feedback on performance. Conversely, systems are weakest in differentiating between level of performance and in providing consequences for performance ratings. Furthermore, half of the agencies that participated in multiple OPM audits made significant improvements to their appraisal systems.

The fact that less than half of the agencies participating in the PAAT process meet the GAO standards raises a number of questions. For example, what percentage of agencies should we reasonably expect to perform well against the GAO criteria, given unique resource environments and political pressures? Whatever our goal should be, it is clear there is room for improvement. Fortunately for agencies, many changes can be made to improve appraisal systems that do not require new legislation from Congress. In some organizations, this will

require designing the changes collaboratively with union representatives as appraisal systems are subject to collective bargaining.

Additionally, what are the substantive consequences of not having an appraisal system that is consistent with the GAO standards? Appraisal systems serve multiple purposes, including documenting performance of other personnel decisions, identify professional development opportunities, providing feedback, and changing behavior and motivating performance (Daley 2004). The idea behind linking the accomplishment of organizational goals in appraisals is based on the assumption that appraisals change and motivate behavior. GAO, OPM, OMB, and politicians assume that performance appraisal systems can motivate and change behavior. However, the analysis presented here does not allow us to test this assertion. Testing this assertion requires scholars to evaluate the relationships between appraisal systems and organizational performance information. Thus far, research has not explored this issue empirically. While many methodological pot holes will need to be navigated in such an effort, it is incumbent upon us to put this assertion to the test.

Information on the structure of performance appraisal systems, beyond perceptual employee survey data, can be used to assess further topics of interest to public management. For example, how do system characteristics relate to issues like satisfaction, commitment, and perceptions of justice? Do variations in structure result in different citizenship behaviors, turnover rates, or complaints? Information on appraisal system structure may also enable scholars to explore questions of equity and goal ambiguity.

For those considering changes to appraisal systems, we have an opportunity to learn from the experiences of the 35 agencies implementing significant changes. Because of the limitations in this dataset, the author was not able to describe these changes in a great level of

detail. However, these 35 can serve as case studies for future research. Examination of these systems can describe the changes made, why the changes were initiated, and the process used for implementation. This would be useful to scholarship on performance appraisals, change management, and organizational performance management more broadly.

Beyond the research implications, this study raises an interesting policy question. Specifically, should agencies be required by law and/or regulation to participate in the PAAT process on a regular basis? OPM currently has no means to require agencies to submit their systems for evaluation. It is clear the OPM standards are consistent with existing legal standards and informed by leading practices. As a result, appraisal systems covering significant portions of federal employees have not been evaluated, and there is no formal inducement to require these units to do so. In the case of Defense, we know the department is re-designing its system after Congress rescinded the reform legislation. However, the situation at Homeland Security is much less clear, particularly in the agencies covered by large union contracts. The value of the data on performance appraisal systems, their evaluation, and the encouragement provided by the evaluation (either via shaming or celebrating) leads the author to suggest that a regular, formal requirement as part of the broader delegated examining unit audit process would be beneficial.

Performance appraisal is a key tool for assuring the accountability of public servants. However, the accountability and performance management literatures often fail to acknowledge this. The information currently available to researchers on appraisal systems in government is derived either from case studies or survey data. Examination of the rules and structures of federal appraisal systems, revealed here, indicates agencies have many opportunities for improvement and scholars have many avenues for future research.

Bibliography

- Ammons, D. and Rodrigues, A. (1986). Performance Appraisal Practices for Upper Management in City governments. *Public Administration Review*, September/October, 460-467.
- Behn, R. (2001). *Rethinking Democratic Accountability*. Washington, D.C.: Brookings Institution.
- Bourgault, J., Dion, S., and Lemay, M. (1993). Creating a Corporate Culture: Lessons from the Canadian Federal Government. *Public Administration Review*, 53(1), 73-80.
- Brewer, G., and Selden, S. (2000). Why Elephants Gallop: Assessing and Predicting Organizational Performance in Federal Agencies. *Journal of Public Administration Research and Theory*, 10(4), 685-711.
- Brewer, G. (2005). In the Eye of the Storm: Frontline Supervisors and Federal Agency Performance. *Journal of Public Administration Research and Theory*, 15, 505-527.
- Brumback, G. (1993). The Continuing Evolution of MBR and Related Developments. *Public Administration Review*, 53(3), 213-219.
- Brumback, G. and McFee, T. (1982). From MBO to MBR. *Public Administration Review*, July/August, 363-371.
- Colby, P. and Ingraham, P. (1982). Individual Motivation and Institutional Changes Under the Senior Executive Service. *Review of Public Personnel Administration*, 2(2), 101-118.
- Daley, D. (2008). The Burden of Dealing with Poor Performers: Wear and Tear on Supervisory Organizational Engagement. *Review of Public Personnel Administration*, 28(1), 44-59.
- Daley, D. (2007). If a Tree Falls in the Forest: The Effect of Grievances on Employee Perceptions of Performance Appraisal, Efficacy, and Job Satisfaction. *Review of Public Personnel Administration*, 27(3), 281-296.
- Daley, R. (2004). Designing Effective Performance Appraisal Systems. In S. Condrey (Ed.) *Handbook of Human Resource Management in Government* (2nd ed.), pgs 499-527.
- Daley, D. (1991). Performance Appraisal in North Carolina Municipalities. *Review of Public Personnel Administration*, 11(3), 32-50.
- Daley, D. (1985). An Examination of the MBO/Performance Standards Approach to Employee Evaluation: Attitudes towards Performance Appraisal in Iowa. *Review of Public Personnel Administration*, 6(1), 11-28.
- deLeon, L. and Ewen, A. (1997). Multi-Source Performance Appraisals: Employee Perceptions of Fairness. *Review of Public Personnel Administration*, 17, 22-36.
- DeVries, D., Morrisson, A., Shullman, S. and Gerlach, M. (1981). *Performance on the Line*. New York: John Wiley and Sons.
- England, R. and Parle, W. (1987). Nonmanagerial Performance Appraisal Practices in Large American Cities. *Public Administration Review*, November/December, 498-504.
- Friedrich, C. (1940). Public Policy and the Nature of Administrative Responsibility. *Public Policy*, 1(1): 3-24
- Finer, F. (1941). Administrative Responsibility in Democratic Government. *Public Administration Review*, 1(4): 335-350.
- Gabris, G. and Ihrke, D. (2000). Improving Employee Acceptance Towards Performance Appraisal and Merit Pay Systems: The Role of Leadership Credibility. *Review of Public Personnel Administration*, Winter, 41-53.

- Getha Taylor, H. (2008). Identifying Collaborative Competencies. *Review of Public Personnel Administration*, 28(2): 103-119.
- Hatry, H. (2007, 2nd ed). *Performance Measurement: Getting Results*. Washington, D.C.: Urban Institute Press.
- Kim, S. (1992). Organizational Support of Career Development and Job Satisfaction: A Case Study of the Nevada Operations Office of the Department of Energy. *Review of Public Personnel Administration*, 22(4), 276-294.
- Lovrich, N., Hopkins, R., Shaffer, P. and Yale, D. (1981). Participative Performance Appraisal Effects Upon Job Satisfaction, Agency Climate, and Work Values: Results of a Quasi Experimental Study in Six Agencies. *Review of Public Personnel Administration*, 1(3), 51-73.
- McNish, L. (1986). A Critical Review of Performance Appraisal at the Federal Level. The Experience of PHS. *Review of Public Personnel Administration*, 7(1), 42-56.
- Mikkelsen, A., Ogaard, T. and Lovrich, N. (1997). Impact of an Integrative Performance Appraisal Experience on Perceptions of Management Quality and Working Environment: Findings from a State Enterprise in Norway. *Review of Public Personnel Administration*, Summer, 82-99.
- Moussavi, F. and Ashbaugh, D. (1995). Perceptual Effects of Participative, Goal-Oriented Performance Appraisal: A Field Study in Public Agencies. *Journal of Public Administration Research and Theory*, 5(3), 331-343.
- Moynihan, D. (2008). *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, D.C.: Georgetown University Press.
- Murphy, K. and Cleveland, J. (1995). *Understanding Performance Appraisal: Social Organizational, and Goal-Based Perspectives*. Thousand Oaks, CA: Sage Publications.
- Nigro, L. (1981). CSRA Performance Appraisals and Merit Pay: Growing Uncertainty in the Federal Workforce. *Public Administration Review*, July/August, 371-375.
- Oh, S. and Lewis, G. (2009). Can Performance Appraisal Systems Inspire Intrinsically Motivated Employees? *Review of Public Personnel Administration*, 29(2), 158-167.
- Page, E., Hood, C. and Lodge, M. (2005). Is Competency Management a Passing Fad? *Public Administration*, 83(4), 853-860.
- Pearce, J. and Perry, J. (1983). Federal Merit Pay: A Longitudinal Analysis. *Public Administration Review*, July-August, 315-325.
- Radin, B. (2006). *Challenging the Performance Movement: Accountability, Complexity and Democratic Values*. Washington, D.C.: Georgetown University Press.
- Romzek, B. S, and M.J. Dubnick (1987). Accountability in the Public Sector: Lessons from the Challenger Disaster. *Public Administration Review*, 47(3): 227 – 238.
- Silverman, S. and Muller, W. (2009). Assessing Performance Management Programs and Policies. In J. Smither and M. London (Eds.), *Performance Management: Putting Research into Action* (pp.527-554). San Francisco, CA: Jossey Bass.
- Van Riper, P. (1958). *The History of the United States Civil Service*. Evanston, IL.: Row, Peterson and Company.
- Weise, D. and Buckley, R. (1998). The Evolution of the Performance Appraisal Process. *Journal of Management History*, 4(3), 233-249.
- U.S. Government Accountability Office (2003). *Results-Oriented Cultures: Creating a Clear Linkage between Individual Performance and Organizational Success*. GAO-03-488.

- U.S. Office of Personnel Management (2010). Federal Employee Viewpoint Survey 2010. Washington, D.C. Accessed May 2011 from http://www.fedview.opm.gov/2010FILES/2010_Govtwide_Report.pdf.
- U.S. Office of Personnel Management (2006). *System Audit Tool: Performance Appraisal Assessment Tool*. Washington, D.C. Accessed in April 2010 from <http://www.opm.gov/perform/PAAT/GS-PAAT-Instructions.pdf>.
- Yang, K. and Kassekert, A. (2010). Linking Management Reform with Employee Job Satisfaction: Evidence from Federal Agencies. *Journal of Public Administration Research and Theory*, 20(2): 413-436.

Table 1
Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max	Number of Agencies Scoring 80% or Greater
Overall Score on GAO Criteria	138	0.665	0.229	0.067	1	60
Alignment	138	0.641	0.337	0.000	1	67
Distinction in Performance Consequences	138	0.550	0.254	0.000	1	18
Employee Involvement	138	0.594	0.222	0.000	1	26
Feedback	138	0.776	0.220	0.000	1	89
	138	0.705	0.306	0.000	1	81

Table 2
Examples of Changes at TSA and the Smithsonian

GAO Variable	TSA			Smithsonian		
	Initial Score	Most Recent Score	Change in Score	Initial Score	Most Recent Score	Change in Score
Overall Score on GAO Criteria	0.22	0.83	0.61	0.17	0.80	0.63
Alignment	0.40	0.90	0.50	0.10	0.85	0.75
Distinctions in Performance	0.09	0.72	0.63	0.09	0.72	0.63
Consequences	0.00	0.60	0.60	0.00	0.60	0.60
Employee Involvement	0.20	0.80	0.60	0.40	0.90	0.50
Feedback	0.20	0.90	0.70	0.17	0.80	0.63
Date PAAT filed	Sept. 2006	Sept. 2009		Jul-06	Jun-08	

Align individual performance expectations with organizational goals. An explicit alignment helps individuals see the connection between their daily activities and organizational goals....

Provide and routinely use performance information to track organizational priorities. Individuals use performance information to manage during the year, identify performance gaps, and pinpoint improvement opportunities....

Use competencies to provide a fuller assessment of performance. Competencies define the skills and supporting behaviors that individuals need to effectively contribute to organizational results....

Link pay to individual and organizational performance. Pay, incentive, and reward systems that link employee knowledge, skills, and contributions to organizational results are based on valid, reliable, and transparent performance management systems with adequate safeguards....

Make meaningful distinctions in performance. Effective performance management systems strive to provide candid and constructive feedback and the necessary objective information and documentation to reward top performers and deal with poor performers....

Involve employees and stakeholders to gain ownership of performance management systems. Early and direct involvement helps increase employees' and stakeholders' understanding and ownership of the system and belief in its fairness.

Figure 1
GAO Criteria for Performance Appraisal Systems
Quoted from GAO (2003)

- **Alignment.** Employee performance plans align with and support organizational goals.
- **Results-focus.** Employee performance plans hold employees accountable for achieving results appropriate to their level of responsibility.
- **Credible Measures.** Employee performance plans provide for balance, so that in addition to measuring expected results, the performance plans include appropriate measures, such as quality, quantity, timeliness, and/or cost-effectiveness, indicators of competencies, and customer perspective. In addition, for managers and supervisors, performance plans should also incorporate employee perspective.
- **Distinctions in levels of performance.** The appraisal program provides for multiple levels to appraise performance and rating officials use those levels to clearly describe distinctive levels of performance and appropriately rate employee performance.
- **Consequences.** The result of appraisal is used for recognizing top performers and addressing poor performance.
- **Employee Involvement.** Employees are involved in the design of the appraisal program and in the development of their performance plans.
- **Feedback and Dialogue.** The appraisal program establishes a performance feedback process that ensures a dialogue between supervisors, managers, and employees throughout the year.
- **Training.** The appraisal program requires that executives, managers, supervisors, and employees receive adequate training and retraining on the performance appraisal program. In addition, supervisors must have the competencies necessary for managing performance.
- **Organizational Assessment and Guidance.** The appraisal program requires that appropriate organizational performance assessments are made and communicated to rating officials, and that guidance is provided by the head of the agency or designee on how to incorporate organizational performance into the assessment process, especially regarding the appraisal of managerial and supervisory employees.
- **Oversight and Accountability.** The head of the agency or designee has oversight of the results of appraisals and awards, ensures that the program operates effectively and efficiently, and ensures that appraisals and awards are based on performance. In addition, managers and supervisors are held accountable for the performance management of their subordinates.

Figure 2
 OPM Criteria for Performance Appraisal Systems
 Quoted from OPM (2006)

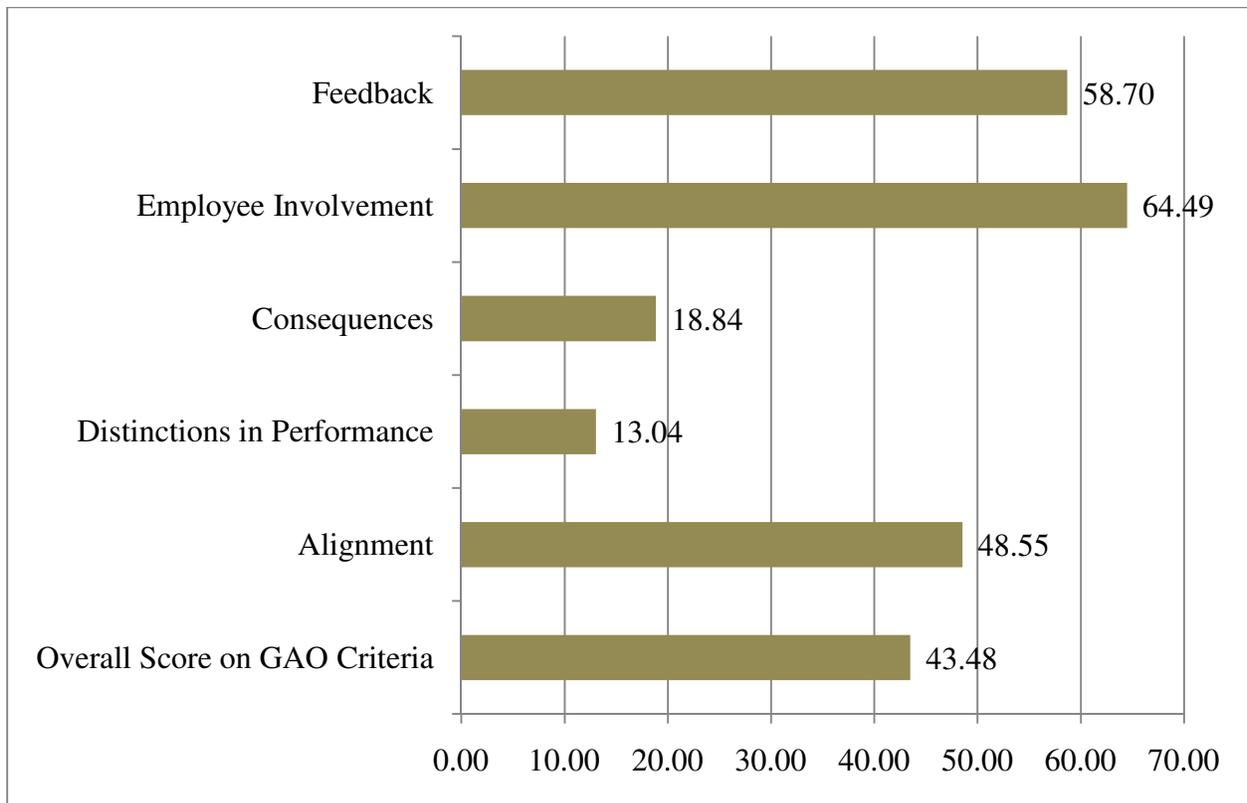


Figure 3
Percent of Agencies Scoring 80 Percent or Greater