Congress of the United States
House of Representatives

COMMITTEE ON OVERSIGHT AND GOVERNMENT REFORM

Subcommittee on the Federal Workforce, Postal Service and

the District of Columbia

Written Testimony Documents

Submitted for Meeting

On

February 12, 2008

By

Dr. Stanley E. Ridley
President & CEO

RIDLEY & ASSOCIATES, LLC

EXEMPLARY HUMAN AND ORGANIZATION DEVELOPMENT CONSULTING

Pay-for-Performance:

Some Observations and Recommendations About the Federal Government's "New"
Approach to Performance Management/Appraisal (PM/PA)[1]

Ultimately, *performance* is the raison d'etre for any organization and its workforce,
regardless of its indicators (e.g., quality services, income, effective regulations, and
satisfied customers). Thus, few miss the obvious irony in the federal government's "new"
approach to PM/PA being called "pay-for-performance." Nonetheless, the goal of having
an effective, efficient and fair/equitable PM/PA system is laudable, and we should
maximize the benefits of this opportunity.

I have more than 30 years of professional experience assessing/testing humans' capability
and functioning, and recommending ways to enhance their performance. My
professional federal government experience extends more than a decade, and has
included providing training and consulting on varied aspects of PM/PA. Some federal
agencies' most recent PM/PA systems include concepts and techniques I developed
and/or recommended. That also appears to be true regarding aspects of how those
systems' are audited, via the Office of Personnel Management's *Performance Appraisal
Assessment Tool*.

Throughout my professional career, I have used scientific reductionism and systems
thinking to isolate the smallest number of factors that explain and address a professional
issue. That approach helps avoid a chronic weakness in many of our problem-solving
efforts: failing to distinguish the problem's *causes* from its *symptoms*. Typically a
problem can have many more symptoms than (underlying) causes. That is definitely true
for PM/PA systems.

My testimony's "cause focus" approach should help explain how to ensure any PM/PA
system works well, regardless of its name. Toward that end, the three key points below
are made relative to a PM/PA system's appropriate conceptualization, design,
development, implementation, measurement, evaluation and periodic refinement.

1. For decades, the requisite technical knowledge, skills and tools (TKSTs) have been
   available, and their potential utility has been significantly enhanced by computer
   technology.

2. For at least a half century, it has been understood that human factors can, and do,
   undermine the effective, efficient and fair/equitable use of the requisite TKSTs.

---

[1] I have explained elsewhere (Ridley, 2007) why PM and PA should not be used
interchangeably. PA should be part of PM. But I will use PM/PA throughout this
document because I am not clear that every federal government agency using a P-F-P
approach clearly makes the needed distinction between PM and PA.

3. The major cause of unsuccessful PM/PA systems is that human basics are allowed to make the requisite TKSTs be underused, misused, or not used.

Much of what I will suggest to address the preceding is reinforced and/or extended in the content of my attached article about pass/fail performance appraisals (Ridley, 2008b).

It is critical that a PM/PA system's decision-makers and other users understand, and appropriately and consistently apply, the information provided below. This must be done **throughout** the conceptualization, design, development, implementation, measurement, evaluation and periodic refinement of any PM/PA system.

The PM/PA system must clearly show that it provides **CARE:**

Clear
Aligned (and)
Reinforced
Expectations

To accomplish the preceding, the **CARE 3-Step** was conceived:

(1) Obtain as much upfront input as possible about content and performance expectations from affected individuals/groups (e.g., managers, staff, stakeholders and customers).

(2) Appropriately incorporate that input, using an *interest-based problem-solving* technique to establish clear and *evidence-based* performance expectations, components and methods that are aligned with the organization's mission, vision, values, strategies and performance goals.

(3) Appropriately and consistently reinforce the performance expectations, components and methods via periodic and constructive feedback, and any other appropriate accountability[2] mechanism.

Three criteria must be used for selecting, using and evaluating any instrument, technique/ method or performance standard in a PM/PA system:

Validity: Degree to which it makes *meaningful differentiation* regarding what it is conceptualized and designed to do.

---

[2] It was not until I started working with the government that I realized the criticality of differentiating accountability from responsibility. I defined responsibility as an obligation/duty, and defined accountability as the appropriate (positive or negative) consequence for how well a responsibility is executed. I also began calling accountability the "Big A" because it became clear that an accountability deficiency was the regnant reason why various government initiatives were not working as well as management, stakeholders, staff and customers would like.

<u>Reliability</u>: Degree of internal consistency in its design, and consistency of its results.

<u>Practicality</u>: Degree to which the resources needed for its development and use are not excessive, especially relative to its benefits.

Those three criteria **must** be used with the following prioritization: validity, reliability, and then, practicality. For PM/PA systems, typically the prioritization is reversed.

PM/PA: It Needs to be a System

The PM/PA components must constitute a system of aligned components. We all recall and understand the importance of this well-worn saying: " a chain is only as strong as its weakest link." Accordingly, Senge (1990) notes that a key characteristic of a system is that its components can be graphically displayed as approximating a circle. This would be evident in a PM/PA system that clearly demonstrates CARE. Too often PM/PA "systems" are really only loosely linked sets of components with destructively diverse degrees of validity, reliability and practicality. Worse yet, some PM/PA "systems" cite key components as being only *suggested*, rather than *required*. As an analogy, think about a car not being required to have a steering wheel or speedometer, and the comparable impact of a PM/PA system that does not have a valid, reliable and practical method for monitoring employee performance throughout the evaluation cycle.

Human Basics and PM/PA Systems

A PM/PA system's success will be clearly linked to how appropriately and consistently it addresses and/or utilizes these human basics:

<u>Motivation</u>: Such as Maslow's needs (safety, security, belonging, etc.), tendency to seek pleasure (e.g., success, being treated fairly) and avoid pain (e.g., negative performance feedback), and being creatures of habit and resisting change; and,

<u>Variability</u>: In capability, functioning, strengths/weaknesses and work-related "styles."

PM/PA System and Decision Makers

To offset the negative impact of human basics <u>and</u> to promote their positive impact, a PM/PA system's success is contingent on a number of appropriate decisions being made during the system's development, use, evaluation and periodic refinement. But decision makers are not monolithic in their views about PM/PA. Consider these views about PM/PA systems from some decision-makers I have encountered:

"I've never really needed it; I know what I, and my people, have to do"

"The one we have might have some weaknesses, but it can't be that bad: look at [the fact that it produced] me"

"They are definitely needed to aid us in developing and rewarding others, and to deal with the atypical employee who has performance problems"

"The only thing they're useful for is when I occasionally come across an employee who does not do what I want/expect"

From the standpoint of a PM/PA *system*, such diverse views are not helpful in decision-making contexts. The tools proposed herein such as the CARE 3-Step would at least reduce this undesirable variability.

PM/PA System Implementation: The Map Does Not Equal the Territory

In the public and private sectors, often an organization's written PM/PA system is at least good—or, it would be with minor tweaking regarding clarity of the performance elements, and ensuring appropriate accountability mechanisms are actually used. Instead, beginning in the 1990s, it became common for organizations to develop "new " PM/PA systems and then, shortly thereafter, drop the "new" system and/or change it markedly. Such occurrences are typically an indication that the changes were largely to address the system's negative symptoms, rather than the underlying cause(s) of those symptoms.

PM/PA System's Deficiencies and Employee Morale and Productivity

Whenever a PM/PA System even appears not to clearly provide **CARE**, perceptions of unfairness (i.e., distributive, procedural and/or interactional injustice) among personnel are likely to ensue. The type and frequency of actual and perceived deficiencies in a PM/PA system's display of CARE, will largely determine its impact on personnel's morale and productivity. As proposed, if a PM/PA system is found to have a significant deficiency, its cause should be readily discernible and amenable to correction in a short period of time.

PM/PA System's Deficiencies and Adverse Impact

Our laws about *adverse impact* in the workplace are indubitably important relative to any PM/PA system. By definition, the demonstration of adverse impact requires a two-step process. First, there must be a statistically significant pass/success rate of the highest scoring/rated employee group, relative to the employee group thought to be experiencing the adverse impact. Second, if that statistically significant difference is found, the organization must show that the instrument(s) and method(s) used to derive the pass/success rates are valid. In the proposed PM/PA approach, the statistical requirement in step #1 would essentially become useless because the fairness to be demonstrated via validity in step# 2 would already have been done as part of the PM/PA systems' development, use and evaluation. Thus, the merit in filing an adverse impact case would be unlikely. The corollary is also true: Failure to ensure the PM/PA system clearly provides CARE increases the likelihood that an adverse impact case could be filed with merit.

Concluding Remarks

Based on the content of this testimony, what do I conclude about how the federal government can ensure that its agencies' "new" pay-for-performance PM/PA systems are successful? Do something that is clearly new. That is, appropriately and consistently provide CARE, by using the CARE 3-Step, and validity, reliability and practicality to develop, implement, evaluate and refine our PM/PA tools so they become systems that are effective, efficient and fair/equitable. And, ensure that pay and other applicable compensation methods are appropriately, consistently and clearly linked to how well the PM/PA system is designed and used by all decision-makers in the respective agencies. The result would be the type of "pay-for-performance" PM/PA systems we all could benefit from, and enthusiastically support!

Bibliography/References

Gladwell, M. (2005). *Blink: The power of thinking without thinking*. New York: Back Bay Books.

Guion, R.M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah: New Jersey.

McGregor, D. (1957). The human side of enterprise. *The Management Review, 46 (11)*, 22-28.

Pulakos, E. D. (2004). *Performance management: A roadmap for developing, implementing and evaluating performance management systems*. Alexandria, VA: SHRM Foundation.

Ridley, S. E. (2007). *Performance management: Needed re-conceptualization and attendant changes to its name, composition, alignment, systematization, measurement and evaluation*. Washington, DC: Ridley & Associates, LLC.

Ridley, S. E. (2008a). *Manager/supervisor performance appraisal learning disabilities: Do you know and prevent the seven deadly types?* Paper accepted for presentation at the Society for Advancement of Management International Business Conference, Arlington, VA, April, 3-6.

Ridley, S. E. (2008b). Measurement and evaluation: OD lessons learned from pass/fail performance appraisals' predictable death. *OD Practitioner*, 40(2), 1-6.

Senge, P.M. (1990). *The fifth discipline*. New York: Doubleday.

Smither, J.W. (Ed.). (1998). *Performance appraisal: State of the art in practice*. San Francisco: Jossey-Bass.

Swan, W. S. (1991). *How to do a superior performance appraisal*. NewYork: Wiley.

"Given organizations' stated performance goals (and measurement and human performance basics), it is inconceivable how they can really have a successful PA process with less than three rating categories for employees' overall performance."

Measurement and Evaluation

OD Lessons Learned from Pass/Fail Performance Appraisals' Predictable Death

By Stanley E. Ridley

Effective measurement and evaluation (M&E) are critical for maximum development and performance of organizations and their employees. The increased demand for effective M&E is traversing the OD landscape at supersonic speed, buttressed by the Zeitgeist, increased workforce diversity and global competition, and varied technical developments such as computer technology, *analytics* and the *balanced scorecard*.

A pillar of an organization's effective M&E should be employee performance appraisal (PA). Unfortunately, PA is in dire need of enhanced effectiveness in most organizations. As this article shows, M&E that actually delivers enhanced effectiveness must be based on solid scientific foundations. "Solutions" without clear evidence of their likely success need to be avoided. For example, the pass/fail approach to PA was hailed as a solution. Seasoned managers, supervisors, human resources (HR) personnel and OD consultants endorsed the P/F approach— not discerning it was doomed to be a failed fad. This article helps explain P/F's rise and predictable failure, and provides OD lessons learned regarding appropriate M&E/PA that can also facilitate the success of varied organizational change initiatives.

The PA Dilemma

Performance appraisal of employees is unquestionably needed, but typically we have difficulty doing them well enough so they do not cause significant problems. That dilemma is reflected in the fact that although formal employee PA processes in the workplace date back at least to the early 1900s, major dissatisfaction with PA's problems was documented consistently during the 1980s and 1990s. For example, Schellhardt (1996) reported that, "In almost every major survey, most employees who get ... evaluations and most supervisors who give them rate the process a resounding failure". That dissatisfaction reached its zenith in the 1990s when exhortations for abandoning PAs received serious debate. Perhaps the voice with the greatest impact in this matter was W. Edwards Deming, the renowned total quality pioneer and guru. His pronounced concern about the American workplace's focus on individuals, rather than on groups and teams as done in Japan, extended to individual PAs (Gabor, 1990). Abandoning formal PAs was often precluded, however, by factors such as union nonsupport, and the government regulations and laws regarding its employees.

Major problems with PAs basically fall within three interrelated categories:
(1) *effectiveness* (e.g., rating accuracy and fairness);
(2) *efficiency* (e.g., time required to complete them), and
(3) *comfortability* (e.g., fear that giving accurate negative feedback could lead to employee grievances).

Pass/Fail Performance Appraisals: Solution or Fad?

Traditionally, most PA approaches have involved using three to five rating levels

(e.g., Above Average, Average, and Below Average). In contrast, the P/F approach only uses two (official) rating levels: Pass and Fail. Reportedly, having only those two rating levels could help address PAs' major problems, key examples including: (1) effectiveness—markedly reducing concern about actual and perceived accuracy of PA ratings, especially among higher functioning employees, and the fairness of linking those PA ratings to personnel decisions; (2) efficiency— markedly curtailing the time required to monitor, record and evaluate employee performance; and (3) comfortability— employees being more receptive to seeking and accepting negative performance feedback from supervisors, and supervisors feeling less anxious about giving it, because such feedback is reportedly de-linked from personnel decisions.

Staunch Support for P/F Appraisals

"Sometimes the solution is worse than the original problem." That saying was prominent in my mind when I first heard about P/F appraisals in the mid-1990s and the staunch backing they were receiving from some seasoned supervisors, managers, HR personnel, and OD consultants I encountered. In contrast, many of those professionals were incredulous when I expressed doubt about whether P/F was definitely the best solution to the PA dilemma. Further, their support for P/F appraisals was unflappable, even though they were unable to provide sufficient evidence regarding such appraisals' effectiveness, and conceptual and measurement soundness.

Actually, through 1995, a major reason for the difficulty obtaining effectiveness information on P/F appraisals was because little existed. By the early 1990s, few organizations had implemented P/F appraisals and, accordingly, even fewer organizations had used them long enough to generate meaningful and reliable success data. Also, access to available P/F appraisal data was a problem. Illustratively, by 1995 only two of the federal government's corporations (and none of its agencies) were piloting P/F appraisals

(Office of Personnel Management, 2000). Further, although a 1994 Hewitt Associates survey suggested 2.4% of private sector organizations were using P/F appraisals, those organizations' promised anonymity as survey participants precluded knowing which to contact and ask about their success using P/F.

By 1995, the limited available data regarding P/F appraisals' success (e.g., Office of Personnel Management, 1995), and anecdotal information, suggested that supervisors tended to like P/F because of efficiency and comfortability benefits such as relative ease of use, time-savings, and avoidance of grievances; but staff had effectiveness/fairness concerns about such fundamentals as how they would be compared with employees not under a P/F approach regarding promotions, awards and reductions-in-force. Related to those concerns, I encountered situations where P/F appraisals had been adopted, and were sometimes already being used, even though the using organization acknowledged that it had not formally determined the alternative method for obtaining the PA information needed to make relevant personnel decisions.

Despite lacking evidence of P/F appraisals' success, and some limited evidence to the contrary, they continued to be hailed and their use increased. For example, in late 1995 the federal government started allowing all of its agencies to use P/F appraisals. By 2000, P/F appraisals were being used with one quarter of its huge employee workforce (U.S. Merit Systems Protection Board, 2001).

Formal evidence of P/F appraisals' non-success was available at least by 1996. Rivenbark (1996) reported that while representatives from three federal government agencies were at a conference touting the process they used to adopt and develop their P/F approach, at the same conference a representative from a government security agency who asked to remain anonymous informed her that the agency was dropping P/F after "several years" of use. The representative reportedly said the reason was that job applicants had no performance information for promotion

boards to review. When the federal government began allowing all its agencies to use P/F appraisals, it discontinued its former one-size-fits-all approach. Perhaps some of its agencies believed that although the P/F approach did not work elsewhere, it still might work in their organizations.

P/F Appraisals: Movement Toward Their Predictable Death

During the 2000s, evidence of P/F appraisals' predicable death has grown exponentially. This is highlighted, for example, by this 2004 Government Accounting Office statement: "We are concerned that a pass/fail system does not provide enough meaningful information and dispersion of ratings to recognize and reward top performance, help everyone attain their maximum potential, and deal with poor performers" (p.1). Accordingly, in 2006 a proposal was considered in the U.S. Congress to no longer allow P/F appraisals in government agencies. Regarding the private sector, I contacted four major organizations that do workplace surveys: Gallup, Society for Human Resource Management, Hewitt, and Watson Wyatt. During the 2000s, P/F's insignificance has been such that neither of those organizations had assessed, or was able to provide me with any data about, the incidence of P/F's usage.

P/F Performance Appraisals: The Autopsy

Why did the P/F approach to performance appraisals fail? The approach was based on two myths, and those myths were ostensibly aided by normal human reactions to the chronic failure of attempts to address the PA dilemma.

Myth #1: Successful Organizations Seeking Maximum Employee Performance Can Have Only Two PA Rating Levels. Not true. To help organizations progress toward maximum performance, personnel attempt to recruit, hire and promote persons most likely to perform at the highest PA level (e.g., Outstanding). Occasionally, limitations such as available salary and benefits may

contribute to organizations settling for persons projected to be only adequate performers. Over time, employees' performance should be expected to vary within at least two levels, such as (1) satisfactory (e.g., Average) and (2) better than satisfactory (e.g., Above Average or Outstanding). Although it is possible that no employee would have functioned less than satisfactorily (e.g., Below Average or Fail), such a category must exist—just in case. Given organizations' stated performance goals (and measurement and human performance basics), it is inconceivable how they can really have a successful PA process with less than three rating categories for employees' overall performance. (Recall in college how unlikely it was for students, maybe including you, to exert maximum effort when taking a pass/fail course.) Organizations can occasionally use two PA ratings levels successfully, however, when applied to components of employees' overall performance (e.g., "violating a key rule or policy").

Further, in practice, the P/F label is probably a misleading misnomer. Even staunch P/F supporters acknowledged that higher/highest level performers (HLPs) exist (such as Outstanding) among those who receive that approach's *pass* rating. But, in the P/F approach, HLPs receive no official PA rating reflecting their performance level. Accordingly, for accuracy purposes, I have suggested that P/F should be considered and called a "2 + 1 tier" (rather than a "2 tier") approach: fail and pass, plus (at least) one HLP level that will be formally determined outside the regular PA process.

Myth #2: P/F Appraisal Ratings Are Really "De-linked" From Personnel Decisions.

They are not. Illustratively, employees cannot be promoted or awarded a raise if they receive a *fail* PA rating. Obviously, however, the impact of pass-fail ratings on personnel decisions is likely to be minimal because so few employees receive the failure rating.

Further, even the purported need to de-link PA ratings from important personnel decisions is specious at best. Why should

personnel decisions not be based on how well employees have performed? If there is a better method to determine the PA rating, use it—in the current PA process (rather than in another process not labeled PA).

The P/F approach's purported de-linking of PA ratings from personnel decisions is really about the HLPs. Perennially, there has been significant concern about the accuracy of HLPs' PA ratings. The P/F approach ostensibly avoids the issue: it provides no official PA rating for HLPs. Responsibility for their PA rating is supposed to be transferred to an alternative rating method (e.g., forced distribution ranking by supervisors). That rating is then considered *linked* to the applicable personnel decision (e.g., annual awards).

The purported P/F appraisal-personnel decision de-linking is most likely to result in *symptom-substitution*. Assume there is an effective alternative method for obtaining PA ratings for HLPs that can be linked to each applicable personnel decision. How would the perennial concern about HLPs' PA rating(s) not just basically transfer to the alternative method? In psychology, decades ago that transfer problem was called *symptom-substitution*: rather than addressing the real "underlying" problem in the original rating method (i.e., inaccurate PA ratings for HLPs), it essentially transfers to the alternative rating method. Subsequently, that alternative rating method, or any of its symptoms, is/are cited as the problem. That occurred with the forced-ranking method for assessing employee performance (cf., Gabor, 1990).

Endorsing P/F Appraisals: Fight-Flight Reaction, Plus Groupthink?

Given P/F appraisal's lethal deficiencies and the talent and experience of many of the approach's staunch supporters, initially I was puzzled about how they could make such an ostensibly illogical endorsement. True, some supporter had only a basic familiarity with, rather than working knowledge of, the basics in measurement—especially the technical

aspects of validity and reliability. That group included seasoned OD consultants. But even when such basics were persuasively presented (such as in the workgroup and training examples described below), P/F supporters wanted to desperately *cling to* that approach. This led me to recall the human *fight-flight* reaction (a response when stress/fear gets too high), and *groupthink* (when group members feel such a strong need to conform that they comply even when the group's stated perception/position is erroneous—resulting in irrational decisions). Accordingly, because many P/F supporters had accumulated numerous negative and stressful experiences with PAs (e.g., angst associated with possible rating inaccuracies and the threat of, as well as actual, grievances and lawsuits), they experienced and presented themselves as basically helpless in their fight to have their organization, colleagues, clients—and themselves—do PAs validly and reliably. In reaction, their ultimate flight would have landed them where PAs were abandoned. Typically, however, as previously explained that was basically impossible. So the farthest they could have taken flight from formal PAs was the P/F approach. Further, groupthink probably buttressed the P/F support. Thus, from this humanistic perspective, the ostensible illogic of such professionals' endorsing the P/F approach essentially disappears.

Addressing the PA Dilemma: Criticality of Applying Basics in Measurement and Human Performance

Given the longstanding PA dilemma and the concomitant need to avoid pre-doomed failures such as P/F appraisals, what can be done to provide real PA solutions? Ensure that such solutions appropriately embody measurement basics: validity, reliability and practicality; and basics in human performance: motivation, capability and variability. Knowledge about those basics is readily obtainable from introductory measurement, psychology, and human development books and courses, as well as the Internet. Cited below are slightly modified versions of

Table 1: Comparison of Pass-Fail and Multi-Level PA Approaches

Criterion (Potential for)	Pass-Fail (Two-Tier/Level)	Multi-Level (Three-Tier/Level or Higher)
Validity (Including Fairness)	Low	Moderate-High
Reliability	High	Low-High
Practicality	High	Low-High

those basics' operational definitions and conceptualizations that I have typically used.

Validity: Degree to which the PAs provide *meaningful differentiation* in employees' performance related to organizational goals and, concomitantly, serve as a key source for determining who should receive special recognition such as awards, bonuses, raises, developmental opportunities, and promotions.

Reliability: Degree of consistency of the PAs' results regarding the likelihood that those results would be similar/the same if different supervisors appraised the same employee.

Practicality: Degree to which the resources needed to complete the PAs are not excessive, especially relative to the PAs' benefits.

Those three measurement basics, I believe, should be the principal criteria for selecting any M&E/PA indicator, and they should be used with the following prioritization: validity, reliability, and then, practicality. My OD experience has been that, if those criteria were used, the organizations' decisions tended to be weighted in the opposite order.

Human Motivation: Few question Abraham Maslow's view that humans are motivated to satisfy their basic needs such as safety/ security. Similarly, it is generally accepted that humans tend to seek pleasure (e.g., success) and avoid pain (e.g., failure and negative feedback), and prefer to be treated fairly.

Human Capability and Performance Variability: For more than a century we have known, scientifically, that human capability and performance vary on a continuum. Often that continuum approximates a normal curve but, depending on the type of attribute, performance and contextual factors, the variability may appropriately deviate markedly from that curve.

Below are two examples of what happened when the preceding measurement and human performance basics were applied to analyses and decisions about P/F appraisals.

Example 1. As an external OD consultant, I was privileged to work with an impressive employee workgroup consisting of management and staff who were charged with helping its organization select a PA approach as part of an overall change management initiative. The organization was having major trust issues, so it was important to foster the workgroup's diversity and, concomitantly, the perceived and actual fairness of the members' selection. Therefore, members were chosen using a stratified random sampling technique, supplemented with self-ratings regarding their degree of interest and availability. (I could not vote on anything the workgroup considered, and its members were to inform management of any behavior/comment on my part that was inconsistent with that, or any of the workgroup's other ground rules.)

An *interest-based problem-solving* technique was used to help the workgroup choose and prioritize criteria for recommending a PA approach. The workgroup selected these prioritized criteria: validity, reliability and practicality. The workgroup also agreed that each member would base his/her vote on which PA approach best fit those criteria— regardless of their personal preference. P/F was called 2-tier (for its two official rating categories) and any other approach was called *multi-tier* (for its three or more official rating categories). Each workgroup

member could share relevant information from any source. The respective pros and cons of each PA approach were presented, and then dialogued in subgroups and by the whole workgroup.

The workgroup was about to vote on which PA approach to recommend when a slight sense of uneasiness surfaced in the room. Then an especially well-respected member brought up some issues—ones that had been previously addressed to the workgroup's satisfaction. Increased uneasiness was pervasive. Then another highly respected member asked if the vote could be delayed to another time/day— without providing any sensible reason for doing so. The uneasiness became more pronounced and the awkward facial expressions and other body gestures suggested the uncomfortability was omnipresent. What caused the preceding? It was a few members' attempted resistance to the inevitable decision—and other members' response to that resistance. Even among members with a staunch preference for P/F, it was obvious that the P/F approach was not recommendable because it did not provide the best fit according to the workgroup's decision-making criteria. Fortunately, the workgroup's development had reached a reasonably high degree of maturity. Accordingly, without solicitation, its members acknowledged that if it had not established the selection criteria beforehand, it would have taken many more meetings to make a decision—if one could have been reached at all. Further, because of the inevitable intra-group dynamics and pressures from colleagues/ supervisors outside the workgroup, they reported not being totally confident that the same decision would have been made.

Table 1 shows the workgroup's final assessment was that although the P/F approach would be highly practical (i.e., little to do) and reliable (most supervisors would agree that all, or almost all, employees passed), it would provide little regarding validity (i.e., meaningful differentiation in performance relative to organizational goals). In contrast, a multi-tier PA approach could have *low to high* practicality and reliability, and *moderate to high* validity.

Example 2. I was facilitating manager and supervisor training sessions on leadership, supervising human resources, and program evaluation. Inevitably, the issue of P/F appraisals arose and needed to be addressed because the participants were actual or potential decision-makers. But P/F appraisals were not a focus of such sessions, so I developed a slightly modified version of *Table 2* below to address the P/F issue in limited time.

Before showing *Table 2*, training participants were asked what percent of their respective workforce was likely to earn a failure rating if a P/F approach were used. With virtually no exceptions (even with my effort to encourage them), the responses were that the failure rate would be one percent or less. After participants agreed to the importance of validity (including fairness), reliability and practicality, and the presented operational definitions of them, they were asked two questions about validity. Q1: "How 'meaningful' is it to distinguish 1% or less of the employees from the other 99%?" Answer: Not much. Q2: "At the beginning of the evaluation year, what percent of employees would think they were going to pass, and what percent of them would actually do so?" Answers to both parts of that question were either 100% or 99–100%. Then they were asked, "How fair is it to treat the performance of everyone in the *pass* category like their performance was the same or comparable?" The answer: Not very fair because those with a *pass* rating are likely to actually fall within more than one actual performance-rating category. *Table 2* shows that it is possible to actually have (at least) five performance groups within the *pass* category, depending on the degree to which they met, exceeded, or did not meet basic expectations.

Many training participants expressed disbelief when they saw that the highly touted P/F approach was so fundamentally weak regarding its validity. Nonetheless, after first acknowledging the veracity of *Table 2*'s information, some training participants still suggested that maybe P/F appraisals should be given "a try." About 1–2 years after those PA training sessions, I was informed that an office in a

Table 2: Analysis of Possible Unfairness of the "Pass" Rating in the P/F Approach

Employees Who "Passed" = ≥ 99%		
Minimum Requirements	**Exceeded Minimum Requirements**	**Possible Ratings**
Passed-Plus 3	Often Much More	Outstanding/Superior
Passed-Plus 2	Occasionally More	Excellent/ Highly Successful
Passed-Plus 1	Rarely More	Satisfactory/Successful
Passed-Barely	Basically Nothing More	Minimally Satisfactory/Successful
Passed—But Should Have Failed	Basically Nothing More— and Sometimes Notably Less	Minimally Satisfactory/Successful
Employees Who "Failed" = ≤ 1 %		
Failed (and Should Have)	Basically Nothing More— and Clearly Notably Less	Unsatisfactory/ Unsuccessful/Failed

government agency was going to *pilot* a P/F approach. That office included supervisors/managers who had participated in at least one of those training sessions.

Summary, OD Lessons Learned/ Recommendations

"If I only knew then what I know now." Despite the importance of M&E and the longstanding PA dilemma, there should have been no opportunity for such Monday morning quarterbacking regarding the P/F approach to performance appraisal (PA). That approach should have been rated *fail* before it initially received a *pass* to be used. The approach's adoption was neither consistent with the prioritized measurement basics of validity, reliability and practicality; nor human basics regarding motivation, and performance capability and variability. The failure of P/F supporters to appropriately consider and use those basics was probably related to a fight-flight reaction, buttressed by groupthink. Moreover, two of the P/F approach's purported key benefits are essentially myths: requiring only two PA rating levels, and those rating levels being de-linked from personnel decisions. Appropriately, P/F appraisals have essentially achieved their predictable death.

So, for OD practitioners what lessons can be gleaned from the grand *faux pas* of the P/F approach to PA? The recommendations below reflect such lessons, and could prove especially valuable because most extend beyond M&E/PA to other organizational change initiatives.

(1) Specifically, for PAs' overall rating levels:
 (a) do not predetermine their number;
 (b) select the largest number that best fits the prioritized selection criteria of validity, reliability and practicality; and,
 (c) if less than three levels are selected, revisit 1a and 1b.
 (The preceding PA recommendations are in addition to those frequently cited basics such as aligning the ratable performance elements with the organization's mission and performance goals, providing individual employee performance plans, and requiring periodic and constructive performance feedback.)
(2) For M&E, PA and Other Change Initiatives:
 (a) OD practitioners should remember to continually use methods that ensure that we have the requisite knowledge, skill and experience to make recommendations, and

work on projects, regarding a given technical area within OD's purview (e.g., validity and reliability of M&E and PA);

(b) relevant human basics (e.g., motivation and capability) should be understood, acknowledged and factored into the process regarding their potentially positive and negative consequences—although this is typically attempted, for example regarding types of *resistance*, too often it is not done as effectively as needed;

(c) select appropriate decision-making criteria *before* deliberations begin about options regarding the change(s) being considered;

(d) make decision-making criteria as transparent and objective as possible via well-crafted operational definitions;

(e) if applicable, appropriately prioritize the decision-making criteria;

(f) have the decision-making team be as diverse as appropriate and possible (e.g., upper and middle management, human resources, supervisors, staff and clients/ customers);

(g) ensure that the decision-making team will be held accountable for adhering to its agreed-upon (prioritized) decision-making criteria;

(h) have at least one expert (external—if the required degree of objectivity is more likely) verify the appropriateness of the operationally defined and prioritized criteria *before* deliberations begin;

(i) have the same/comparable expert(s) verify the degree to which each recommended decision aligns with the decision-making criteria; and,

(j) establish valid and reliable mechanisms for accountability during and after the change process (that *actually* hold the applicable people accountable).

If the preceding recommendations are followed, not only can an organization's

M&E, PA and other change initiatives be more likely to succeed, their success level will probably merit an award. So be sure to have predetermined and fair M&E criteria for that purpose too—that go beyond merely asking if the change effort passed or failed.

References

Gabor, A. (1990). *The man who discovered quality.* New York: Times Books.

Government Accounting Office (2004, September 3). *Posthearing questions related to assessing progress in human capital management* (Abstract). Washington, DC: Author.

Hewitt Associates. (1994). *The impact of performance management on organizational success.* Lincolnshire, Illinois: Author.

Office of Personnel Management. (October, 1995). Pass/fail appraisals. *Workforce Performance Resources Newsletter.* Washington, DC: Author.

Rivenbark, L. (1996, December 9). Grade is still out on pass-fail appraisals. *Federal Times.* [Page number(s) unavailable.]

Schellhardt, T. (1996, November 19). Annual agony: Its time to evaluate your work and all involved are groaning. *Wall Street Journal,* p. A1.

U.S. Merit Systems Protection Board. (2001, August). Performance appraisals are not performance management. *Issues of Merit,* p.1.

Stanley E. Ridley, PhD, is President & CEO of Ridley & Associates, LLC, an organization development consulting firm located in Washington, DC. His more than 30 years of professional experience has included roles as a tenured university professor, trainer, researcher, administrator, supervisor, executive coach, organization and human development consultant, clinical psychologist, and test/assessment developer and evaluator. Most recently, he developed state-of-the-art performance appraisal and performance management systems. Stan earned his doctorate degree in psychology from Howard University. He can be reached at ser@ridleyassociatesllc.com.