ANNUAL REVIEWS

Annual Review of Organizational Psychology and Organizational Behavior

The Evolution of Performance Management: Searching for Value

Elaine D. Pulakos,[1] Rose Mueller-Hanson,[2] and Sharon Arad[3]

[1]PDRI, Washington, DC 22209, USA; email: elaine.pulakos@pdri.com

[2]Community Interface Services, Carlsbad, California 92054, USA; email: rose.hanson@gmail.com

[3]PDRI, Minneapolis, Minnesota, 55401, USA; email: sharon.arad@pdri.com

Annu. Rev. Organ. Psychol. Organ. Behav. 2019. 6:249–71

First published as a Review in Advance on July 13, 2018

The *Annual Review of Organizational Psychology and Organizational Behavior* is online at orgpsych.annualreviews.org

https://doi.org/10.1146/annurev-orgpsych-012218-015009

Copyright © 2019 by Annual Reviews. All rights reserved

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

performance ratings, performance management processes, PM streamlining, behavior change, PM innovation, PM implementation

Abstract

This article reviews the history of performance management (PM), beginning with performance evaluation. We discuss various strategies that have been used to enable accurate ratings as well as cognitive processes and contextual factors that have been shown to significantly impact ratings. We raise questions about the concept of true performance and whether raters can be enabled and motivated to make accurate ratings. We progress to discussing more structured and comprehensive PM processes that typically involve cascading goals, goal setting, competency modeling, evaluation of behavior and results, and implementation. These systems have proven to be tedious and low-value, resulting in unprecedented, wide-spread experimentation with innovative practices to move companies away from heavy PM processes to simpler, cost-effective strategies that actually drive performance. These have ranged from abandoning ratings to implementing innovations in goal-setting, real-time feedback, coaching, and PM behavior change. Directions for future research and practice are discussed.

INTRODUCTION

No other talent management system has been the subject of such great debate, change, and emotion as performance management (PM). Many strategies have been attempted to extract value from PM processes, ranging from simple rating scale changes to complex behavior change initiatives. Although many of these have seemed promising initially, they have yielded disappointing outcomes and significant dissatisfaction with PM processes in organizations. The challenges inherent in PM coupled with many unsuccessful attempts to fix it have given rise to what are viewed as trendy PM practices and a well-earned reputation as the Achilles' heel of talent management practice (Pulakos et al. 2012). No matter what has been tried over decades to improve PM processes, they continue to generate inaccurate information and do virtually nothing to drive performance.

PM is challenging because it is a complex, multifaceted, and multilevel process that draws on theory and research from many different areas, including measurement theory and motivation theory; cognitive, clinical, social, and behavioral psychology; neuroscience; organizational development; and change management. It has been heavily influenced by practice as well, with business leaders and PM practitioners offering rating schemes and evaluation strategies they believe will drive higher performance. Research has generally focused on specific aspects of the PM process, such as the effects of different types of rating scales on ratings or the role of human information processing in evaluating others (DeNisi & Murphy 2017). Only a few studies have focused on the impact of different PM practices on performance and which features play the greatest role in driving business outcomes; thus, we have relatively few evidence-based insights about the impact and return on investment (ROI) of PM in organizations.

Over time, PM has become increasingly complex, requiring many hours of manager and employee time and costing organizations millions annually. CEB (2012) estimated that the average manager and employee spend 210 and 40 hours, respectively, on PM activities, which translated into costs of 30 million USD annually for a company of 10,000 people. As another example, Deloitte found that they were spending 2 million hours annually on PM activities (Buckingham & Goodall 2015). Not only have time investments and costs skyrocketed, but complaints have become increasingly vocal and emotional, especially concerning performance reviews. Culbert & Rout (2010) described the performance review as a "pretentious, bogus practice" that should be put out of its misery. A *Washington Post* headline read, "Study finds that basically every single person hates performance reviews" (**https://www.washingtonpost.com/news/on-leadership/wp/2014/01/27/study-finds-that-basically-every-single-person-hates-performance-reviews/?utm_term=.03749c8f8b7d**).

Most concerning, however, is that the few studies that have evaluated the impact of PM processes on performance and business outcomes have shown virtually no positive impacts. For example, based on 23,339 performance ratings from 40 organizations, CEB (2012) found that business units with highly rated employees were no more likely to be profitable than those with low-rated employees. So much cost and time, yielding so much dissatisfaction with no discernable performance impact, has led to a recent wave of sweeping, revolutionary reform, and experimentation with new-in-kind PM practices. These range from greatly simplifying or even eliminating formal PM processes to driving behavior changes that research has shown positively impact performance, such as providing regular informal feedback and setting agile, shorter-term goals (Mueller-Hanson & Pulakos 2018).

In this article, we briefly review the evolution of PM, which began with a much narrower focus on performance ratings. We trace its development from research evaluating the impact of rating format and training on ratings to that aimed at understanding how human information processing, rater-ratee interpersonal relationships, and political and contextual factors affect ratings. We then

discuss how performance evaluation evolved into more comprehensive PM processes that included goal-setting, formal feedback, multirater reviews, etc., and we discuss how these practices have become misguided over time. Finally, we discuss current directions in PM, taking stock of what we have learned to date and suggesting directions for the future.

PERFORMANCE EVALUATION

The early history of PM was focused on performance evaluation, the goal of which was to obtain accurate ratings of individual performance. The first large-scale use of ratings in work settings dates back to the late 1800s, with use of efficiency ratings in the US Federal Civil Service (Lopez 1968) and trait assessments (e.g., punctual, assertive) of officer performance during World War I (Scott et al. 1941). The first rating scale, the Graphic Rating Scale (Patterson 1922), used verbal and numerical anchors to improve the accuracy of trait ratings. Although this was a significant step, the anchors used were ill-defined (e.g., "Excellent," "Good," or "Poor"), leaving raters to impose their own interpretations on what these anchors meant (Landy & Farr 1980, Borman 1977). Raters applying their own idiosyncratic standards to defining different rating levels remains a persistent challenge today.

The emergence of scientific management theories in the early twentieth century (Taylor 1911) led to an increased focus on productivity and the corresponding use of ratings to control and drive higher performance (Grote 1996, Murphy & Cleveland 1995). The civil rights movement of the 1950s and 1960s brought attention to inequalities based on race and prompted more rigorous evaluation practices in organizations. The Civil Rights Act of 1964 and subsequent legislation prohibited discrimination in employment practices, prompting extensive work in the area of rating format design to ensure ratings were based on job-relevant factors and to mitigate bias (Dunnette 1963, Guion 1961). One idea that gained popularity was to anchor different rating levels with work behaviors to help managers match their observations of employee performance to an appropriate rating level (Smith & Kendall 1963, Blanz & Ghiselli 1972, Latham & Wexley 1977). Many variants of behavioral rating formats were designed and evaluated over the next 20 years, until Landy & Farr (1980) called for a moratorium on rating format research, concluding that no rating format yielded substantially more accurate or less biased ratings than any others (Murphy et al. 1982, Saal & Landy 1977).

Although rating format research largely ceased with Landy & Farr's (1980) moratorium, a new forced choice rating format was introduced in the early 2000s that has been shown to yield improved rating reliability, validity, and accuracy (Borman et al. 2001, Bartram 2007, Schneider et al. 2003). This format asks managers to choose which behavior is most true (or most and least true) of each employee's job performance from a set of equally desirable behaviors. Using item response theory (IRT) information for each item, raters' judgments are converted to an interval scale; specifically, choosing one behavioral statement over the others provides information about the placement of each employee on the underlying dimension at the interval-scale level. Although research has shown this format to yield higher quality ratings, its adoption has been rare in practice. One reason is that advanced IRT concepts are difficult to explain. Forced choice formats also require large item banks with associated item parameters that can be prohibitive for organizations to develop and maintain. Finally, the main advantage of forced choice ratings is also likely its main disadvantage, namely that managers cannot easily manipulate their ratings to ensure employees receive certain reward outcomes; hence, forced-choice scales are not well received by managers.

A parallel path to improve ratings focused on rater training (Borman 1975, Latham et al. 1975). On the basis of the assumption that ratings are normally distributed, training programs were developed to teach raters to avoid common rating errors that would result in non-normally

distributed ratings, such as leniency (most employees are rated at the high end of the scale). To reduce leniency, for example, raters were taught that most employees should be rated in the middle of the scale and equal but smaller proportions should be rated at each of the high and low ends. Subsequent research showed that error training did not increase accuracy and may actually reduce it (Murphy et al. 1993). Years later, O'Boyle & Aguinis (2012) provided evidence that performance is not normally distributed in many cases, which explains why training to produce normally distributed ratings would decrease accuracy.

A paradigm shift occurred in the early 1980s that influenced performance rating research for the next two decades. Landy & Farr (1980) argued that more holistic theories were needed to understand the interactive effects of different factors on ratings, and they proposed the use of human information-processing theories and models to guide future research. Ratings were conceptualized as a special case of human information processing that includes attention, categorization, recall, and information integration (Feldman 1981). Extensive research leveraged information-processing theories to understand rating behavior and develop interventions to improve rating accuracy. These focused on helping raters develop and use job-relevant mental categories in observing and evaluating employee performance (Ilgen & Feldman 1983). For example, rater training shifted from a focus on reducing rating errors (e.g., halo, leniency) (Cooper 1981, Murphy & Balzer 1989, Murphy et al. 1993) to helping raters create job-relevant mental categories that would direct their attention to relevant performance information and store it with related performance information to facilitate accurate recall (McIntyre et al. 1984; Pulakos 1984, 1986). Although information processing theories provided insights into the mental processes that impact how ratings are made, this research yielded few practical implications for evaluating performance more effectively in organizations.

Mounting concerns over discrimination and legal challenges in the 1970s and 1980s brought implementation of more structured evaluation processes. For example, management by objectives (MBO; Drucker 1954) provided a way to define, communicate, and evaluate employees against job-relevant performance objectives. Although MBO systems were widely adopted, they were eventually abandoned because they proved to be time-consuming and administratively burdensome for their value (Jamieson 1973, Strauss 1972). However, ideas stemming from MBO, such as setting objectives and measuring results, remain a common feature in PM processes today.

A popular rating method to emerge in the early 1980s was the forced distribution, introduced by former General Electric CEO Jack Welch. Known as GE's "rank and yank" system, employees were slotted into categories based on how their performance stacked up to other employees', with small proportions (10–15%) identified as top and bottom performers and the remaining ~80% slotted in the middle. The top and bottom groups often defined those to be promoted and separated, respectively. The practical problem forced distributions posed is that the top 10% in a low performing group may be performing with the same effectiveness as the bottom 10% in a high performing group, introducing both fairness and accuracy concerns if the groups are blindly combined. This issue is typically mitigated through calibration sessions in which employees are discussed and recategorized to ensure that the top and bottom 10% are accurately identified across all employees. However, this is a time-consuming process that becomes less informed as calibration rolls up through higher organizational levels and individual employee performance at lower levels becomes less well known. Although forced rankings remained popular for more than 30 years, their use is now on the decline, falling from 49% in 2009 to 14% in 2011—GE being among those to abandon this rating method (i4cp 2011).

Another rating strategy that emerged about this same time was gathering multisource or 360-degree ratings from peers, customers, or direct reports in addition to managers. The idea was that those with different role relationships to an employee observe different aspects of

performance (Borman 1974). For example, customers will have unique insights into one's customer service effectiveness, whereas direct reports will be best equipped to evaluate a manager's feedback and mentoring performance. 360-degree ratings gained popularity in the 1980s and are still widely used today (Bracken et al. 2001, Smither et al. 2005). They are primarily used to provide developmental feedback, but they can also support decision making, if the rating information from the different sources is appropriately integrated and interpreted by the person's manager or a coach (Bracken et al. 2001). One caveat, however, is that decrements in the quality of multisource ratings are often observed when they are used for decision making versus development only (Greguras et al. 2003).

Performance Evaluation Challenges

Underlying performance rating research are three assumptions that are worthy of further exploration:

- Everyone has a stable level of true performance that reflects their effectiveness on the job.
- Raters are able to rate others accurately.
- Raters are motivated to rate others accurately.

Regarding the first, we assume that each individual has a "true" level of performance that they consistently exhibit on the job. We then use the extent to which different raters agree on their ratings of an individual as an indicator of how well the ratings are capturing a person's true performance level, with higher agreement giving us more confidence we are accurately measuring the person's true performance level. Rater agreement is hard to achieve, however, in part because raters bring their own standards to any rating situation, based on their past experience, personal rating tendencies, and idiosyncratic views about what constitutes good or poor performance (Landy & Farr 1980, Feldman 1981). However, rating disagreement can also stem from raters viewing different aspects of performance or, importantly, real differences in how employees actually behave in the presence of different raters. An individual may be highly responsive with managers but disregard peers—or the person may help only some peers but not others. These realities explain why interrater reliabilities are typically only in the .50 range (Viswesvaran et al. 1996), and they also raise questions about the extent to which true performance can be agreed upon among raters, or even exists.

The second assumption is that raters can make accurate ratings with proper rating instruments and training. The reality is that most managers can identify who is doing a job capably, who is failing, and who is performing above and beyond. However, the ratings managers are asked to make are sometimes so nuanced and detailed that they are beyond their information-processing capabilities. Managers see thousands of performance examples in a year-long rating cycle—far too many to recall, weight, and summarize with a high degree of accuracy for each employee. Furthermore, managers do not see performance in some areas (e.g., how employees engage with their direct reports) and may not have the subject matter expertise to judge some of what they do see (e.g., general managers rating highly technical performance), causing them to rely on biased impressions, what others say, or stand-out examples of obviously exceptional or poor performance (Landy & Farr 1980). Rating scales that contain many rating levels or factors that require highly nuanced judgments are asking for rating precision that managers cannot realistically provide (Pulakos & O'Leary 2010). Many overengineered rating formats and processes have been developed that do not align well with raters' information-processing capabilities.

The third assumption is that raters are motivated to evaluate others accurately. However, several studies question this assumption by showing that various contextual factors undermine rating

accuracy (Tziner & Murphy 1999, Murphy et al. 2004). Murphy & Cleveland (1995) suggested four competing goals that managers must negotiate and balance when they evaluate employees:

- Task performance goals, which entail using ratings to influence subsequent performance.
- Interpersonal goals, which entail using ratings to maintain or improve relationships with employees.
- Strategic goals, which entail using ratings to increase the manager's or workgroup's standing in the organization.
- Internalized goals, which reflect raters' personal beliefs about how they should evaluate performance.

It has been proposed that political, social, and practical factors carry so much weight in managers' rating behavior that rating accuracy and employee differentiation are simply not relevant drivers of ratings (Adler et al. 2016). It has similarly been argued that managers have few if any incentives to rate employees accurately (Pulakos & O'Leary 2011). Unless an employee is a problem performer, many managers take the pragmatic approach of playing to employees' strengths and assigning them work they can do well (Mueller-Hanson & Pulakos 2018). They realize that all employees are imperfect and each brings different capabilities to a job. If employees—especially experienced employees—are making solid contributions, managers often overlook weaker areas for which there is little chance of change or growth. Experienced managers understand the practical realities and costs of replacing staff and onboarding new staff that will also be imperfect. Finally, they understand that employees want to be recognized and praised, which creates a strong incentive for them to rate their key, albeit imperfect, employees above the midpoint of any scale in order to keep them motivated and engaged.

Managers also take a pragmatic approach to how they use ratings in pay and reward decisions (Pulakos et al. 2015). Instead of using the rating process to arrive at an evaluation and then translate this into a pay decision, managers are more likely to retrofit their ratings to align with the reward decisions they want to make at a given point in time. The practical considerations that drive pay decisions include mitigating attrition risk, managing internal or external equity, and even whose turn it is to get a larger increase—a phenomenon that results from the relatively small (2–3%) raise pools that most organizations have today. To the extent that ratings align with pay increases, it is often the latter driving the former.

Given that a manager's job is to get the highest performance out of his or her team, a key part of the job is using all available levers to keep the collective group engaged, productive, and performing. Managers also have their own motivations and advancement goals that the perception of a high performing and engaged team helps them achieve. Although rating everyone at the high end of a rating scale may not yield accurate evaluations, it can be argued that this is rational behavior, especially when today's managers are being asked to do more with less, they want their teams and themselves to look good, and they want access to rewards and future opportunities (Mueller-Hanson & Pulakos 2018). Context factors thus have profound impacts on ratings and their implications need to be better understood and accounted for in the design of PM processes.

Summary and Next Steps for Performance Evaluation Research and Practice

What we have learned over decades of research and practice is that performance ratings bring significant challenges. Employees behave differently with different raters due to role and relationship differences. Managers bring their own standards, levels of sophistication, and expertise to evaluating others. They are swayed by their own biases, differences in the quality of their relationships

SUMMARY: PERFORMANCE EVALUATION

Below is a summary of our current conclusions regarding performance evaluation based on research and practice to date:

- Ratings are inherently limited in their value as performance measures.
- Rater-ratee relationship differences yield actual performance differences, which raises questions about whether a "true" performance level exists that can be reliably captured across raters.
- Raters can accurately place others into general categories but cannot make nuanced performance judgments accurately.
- Political and social factors have very strong impacts on ratings.
- Properly selected, performance measures beyond ratings may mitigate challenges with ratings.

with different employees, and their rating preferences. They also see only a slice of each person's work behavior and may or may not have the expertise to accurately evaluate what they see. These factors make true performance impossible to define and rating accuracy impossible to evaluate. The cognitive processes humans naturally use to process performance information leave raters with summarized impressions rather than detailed performance information. Although heavy requirements for precise and nuanced ratings may inspire more confidence that we are closer to ground truth about an employee's performance, this is a false sense of confidence, because there is no evidence that more complex ratings improve accuracy or fairness. An important consideration is how much rating differentiation and accuracy is actually needed, especially when ratings have been shown to have no impact on performance. Unless highly nuanced differentiation is required to distribute significant rewards, complex rating processes are unlikely to yield ROI commensurate with their costs. Simpler judgments that align with the overall judgments raters naturally make are likely sufficient and most practical for the majority of evaluation needs (see sidebar Summary: Performance Evaluation).

Given the challenges inherent in ratings, performance evaluation can benefit from leveraging measures other than ratings. In some jobs, a great deal of performance information is readily available beyond ratings, such as customer surveys, sales data, production data, efficiency indices, and billable hours. Although these measures also have limitations (e.g., they can be deficient or contaminated), they can provide a more well-rounded picture of performance that goes beyond ratings. When collected on a regular basis, such measures can be used to signal performance issues early and drive real-time feedback to course-correct. An added benefit is that nonrating measures lessen the pressure on managers as their role can shift from judging employees to helping them understand and respond to different performance measures. The use of multisource ratings combined with attention to performance measures that exist in the environment may reduce the impact of political and social factors on ratings. With the level of digital transformation that is occurring in organizations coupled with the increasing focus on analytics, we will see increasing availability of and focus on performance measures that are automatically and frequently generated. The questions for future performance evaluation research are (*a*) how to leverage and combine available performance information (e.g., various metrics, ratings, etc.) data into meaningful, sensible, valid, and fair performance assessments and (*b*) what role humans will play in future performance evaluation processes. The answers to these questions will almost certainly result in new performance measurement strategies and practices that may mitigate the limitations of ratings but will also need to be carefully evaluated for their potential consequences.

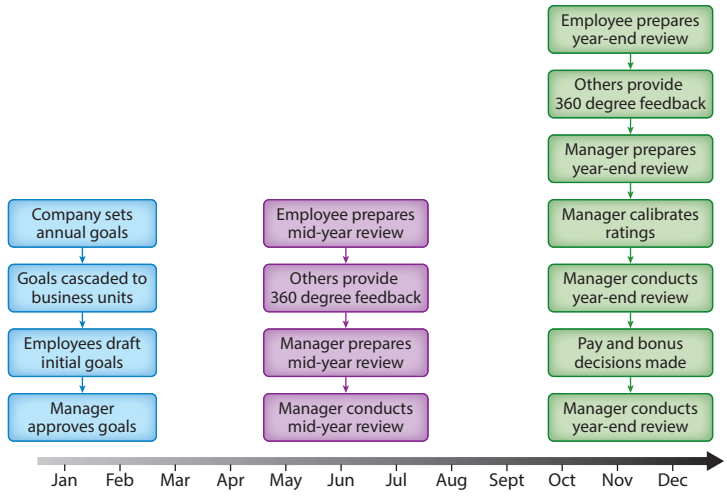

Figure 1

Typical performance management process used in organizations.

PERFORMANCE MANAGEMENT

With flatter, leaner organizations and pressure to do more with less, performance evaluation eventually evolved into more comprehensive PM processes that included a fuller array of activities to drive performance, such as cascading goals, expectation setting, and interim feedback reviews (Smither & London 2009, London & Mone 2014). These processes became fairly standard over the past 15–20 years, especially as organizations began acquiring automated PM systems to improve its efficiency (Aguinis 2013, London & Mone 2014). In these systems, employees are usually evaluated on behavior and results (Pulakos 2009). The idea is that both "how" employees perform (behaviors) and "what" they deliver (results) are considered important aspects of their performance. Behavioral ratings are more useful for course-correcting than results, which come after the fact. Behavioral ratings are notoriously attenuated, however, with most people rated above the midpoint of the rating scale, reducing their usefulness (Pulakos 2009). Results capture what some argue is most important—the outcomes one achieves—although results measures can break down when goal attainment is outside an employee's control or results from team performance rather than individual performance, which is often the case today (Locke & Latham 1990, Ployhart et al. 2009). A typical automated PM process is shown in **Figure 1**.

Most PM processes begin by setting goals and objectives for each employee—a practice rooted in goal-setting research, which shows that employees perform more effectively when specific goals are set (e.g., Locke et al. 1981, Locke & Latham 1990). Cascading goals are often used to link the organization's strategic goals down to each employee (Rodgers & Hunter 1991). The idea is that these linkages will help employees understand how their work aligns with the organization's strategy and goals (Hillgren & Cheatham 2000, Schneier et al. 1991). Leveraging principles from MBO, objectives state the outcomes each employee is expected to achieve in sufficient detail to judge whether the objective has been met. Employees and managers are often trained to set SMART (Specific, Measurable, Aligned, Realistic, and Time-Bound) goals as part of this process.

Although the idea of linking individual and organizational goals makes sense, Pulakos & O'Leary (2010) noted the following practical challenges:

- Cascading goals take time and can be difficult for managers who are not accustomed to linking goals between levels.
- As goals are cascaded, they often become disconnected from organizational goals and obscured, akin to the game of telephone, in which retelling a story can alter it in ways not intended.
- Even with training, the quality of the objectives varies greatly from manager to manager, and objectives are rarely comparable across similarly situated employees (Pulakos & O'Leary 2010).
- Even when jobs are predictable, goals set at the beginning of the year cannot account for unexpected events during the year. This challenge is exacerbated in fluid situations in which priorities change frequently (Cascio 1998, Pulakos & O'Leary 2010). Although guidance is given to update objectives as the situation changes, this is rarely done in practice.
- It can be difficult to assess the relative contributions different employees make when their objectives are not comparable (Pulakos 2009). Achieving an easier goal fully may yield much less contribution than partial delivery of a challenging goal, yet the former will typically be rated higher than the latter.
- Finally, goal attainment is often based on available rather than optimal measures, which can sacrifice important criteria, for example, measuring quantity rather than quality.

Although extensive research has been devoted to goal setting, most has been conducted in laboratory rather than field settings, leaving several questions unanswered in work situations. For example, little research has evaluated the extent to which measures of goal achievement are reliable, valid, and fair measures of performance (O'Leary & Pulakos 2017). Questions have also been raised about the motivational impacts of goal setting over long time horizons, with research showing that shorter time spans are better for driving performance (Latham & Locke 2007). However, performance objectives typically cover year-long timeframes, which are likely too long to maintain strong motivational effects. Finally, rewarding goal attainment can change its dynamics dramatically. If employees know they will be rewarded by achieving a goal, they will angle for highly attainable goals to guarantee rewards, which is why goals in performance evaluation contexts tend to be less aspirational and challenging than goals in learning contexts (Winters & Latham 2006). The above factors raise questions about the extent to which goal setting in a PM context as it is accomplished today is actually useful in driving high performance.

A common feature of PM processes is the use of competency models as the basis for behavioral ratings. David McClelland is credited with introducing the idea of a competency (Dubois 1993), and competency-based talent management practices have developed rapidly ever since, although there has been significant debate about what exactly competencies are and how to most effectively measure them, with some critical views. Boyatzis (1996) defined a competency as a combination of a motive, trait, skill, attribute, or a body of relevant knowledge; in other words, a competency is any individual characteristic that is related to successful job performance. Klein (1996) differentiated competencies from psychological constructs, defining them as related, observable behaviors representing common themes that differentiate effective from ineffective performance. Competency models today typically consist of several behaviorally defined performance factors, following Klein's concept. Competency models provide a useful mechanism to articulate an organization's strategy, values, culture, and priorities. Over the years, more rigorous approaches to competency modeling have evolved that have a job-analytic backbone to better support their use in evaluation and decision making (Schippmann 1999). In a PM context, ratings are made on competencies that are relevant for each employee's role. However, similar to the rating format research discussed above, competency-based rating formats have proven to be no more accurate or unbiased than any other rating formats.

In spite of considerable investment in designing and automating PM processes, automation did not yield improvements in PM quality, effectiveness, or impact, which prompted increased attention to implementation activities such as building leadership support and employee buy-in and developing effective communication, training, and change management strategies (Rodgers et al. 1993). Large-scale survey research studies began examining the impact of PM processes on employee engagement and performance (CEB 2004, Harter et al. 2002). Concerns raised by this research, which showed PM processes to have little if any impact on performance, led initially to the development of more elaborate processes. PM steps were broken down into more specific, structured, and scheduled PM activities, based on the idea that more prescribed steps and time devoted to PM would translate into improved outcomes. This resulted in more burdensome PM requirements for managers and employees that became increasingly disconnected from what managers actually need to do to manage performance on a daily basis, such as ensuring employees know what they need to do and coaching them to overcome barriers in real time. Pulakos & O'Leary (2011) have argued that automation has exacerbated the schism between formal PM and how managers manage performance day to day, by turning an inherently human and fluid process into intermittent, automation-enabled administrative steps.

Negative attitudes toward PM, especially formal review sessions, reached fever pitch in the late 2000s, with emotional calls to eliminate them (Culbert & Rout 2010, Culbertson et al. 2013). Neuroscience research provided compelling evidence of brain changes that initiate automatic defensive reactions to formal performance reviews, showing physiological mechanisms that make these aversive for even high performers (Rock 2008, Rock & Jones 2015). Further fueling negative affect toward formal PM processes were ROI analyses showing millions of dollars in costs and excessive time devoted to PM activities that are uniformly viewed as low value and have no impact on individual or organizational performance (CEB 2012). The combination of these factors has led to a massive movement to reinvent PM. There has never been a time when so many companies have experimented with disruptive change to a major talent system, which speaks volumes about the dissatisfaction with PM (Adler et al. 2016).

New Approaches to Performance Management

The past five years have brought novel ideas for how to gain more value from PM that are different in kind from traditional practices. Rather than continue to add more formal PM steps and processes, an overwhelming number of companies have begun stripping down their formal processes and focusing more on informal, unscheduled behaviors that have been shown to matter most in driving performance (CEB 2004, Bryant 2011). A plethora of case studies have been published over the past few years describing how different companies approached their PM reform. For example, Deloitte's PM transformation was highlighted in the *Harvard Business Review* (Buckingham & Goodall 2015) as well as numerous online articles (Baldassarre & Finken 2015, Mosley 2015, Rock & Jones 2015). Interest has been so high that mainstream media outlets such as the *New York Times* and *Washington Post* (Cunningham 2015, Kantor & Streitfeld 2015) have featured articles about PM reform at Accenture, Amazon, Deloitte, General Electric, and Microsoft, among others.

The number of companies working on PM reform simultaneously has generated considerable debate about which changes will add the most value, resulting in two primary approaches. The first approach is to change aspects of the formal system to reduce unnecessary, low-value steps and processes that yield high costs with no clear ROI. This strategy to fix PM aligns with general trends to reduce talent management process complexity, which Effron & Ott (2010) argue has become too heavy, burdensome, and costly for its value. The second major approach to improve PM focuses

on the day-to-day manager and employee behaviors that have been shown to drive performance, such as providing real-time feedback that helps employees perform better or overcome challenges and implementing agile goals that are nearer term and easier to adjust as the situation changes. Some organizations have focused their efforts on the first strategy (streamline the process), others have focused on the second strategy (drive more effective PM behavior), and some have made changes that leverage both approaches in tandem.

Streamlining the Formal Performance Management System

The design of new PM processes needs to start with careful consideration of the purpose(s) for PM, which can include development, administrative (e.g., decision making around pay, bonuses, promotion, etc.), or other (e.g., legal defensibility) goals. The primary purpose(s) of PM in a given organization must be clearly defined, because PM processes that try to serve too many purposes often end up serving none well (Pulakos 2009). Clarity of purpose helps in two key ways:

- It enables organizations to avoid PM goals that fight each other; for example, a longstanding finding is that evaluation goals can interfere with effective development. When rewards are in the mix, employees are reluctant to focus on development and ratings become more corrupt, which we discuss in more detail below.
- It provides an important benchmark against which potential PM features can be evaluated to avoid including peripheral features and complexity that do not add incremental value.

Although organizations have experimented with streamlining their goal-setting processes, formal reviews, and rating process, they have become most fixated on whether to retain ratings, which has become the subject of considerable debate (e.g., Adler et al. 2016). Many high-profile companies (e.g., Accenture, Deloitte, Microsoft, GAP) have abandoned or substantially reduced their use of ratings (Culbert & Rout 2010, Cunningham 2015), whereas others have held firm that ratings are important to keep. If the debate about ratings concerned questions about rating quality or accuracy, the answer would be simple. Nearly a century of research shows longstanding problems with ratings, including inaccuracy, questionable fairness, and low value (Bernardin & Beatty 1984; DeCotiis & Petit 1978; DeNisi 2006; Ilgen & Feldman 1983; Landy & Farr 1983; Murphy & Cleveland 1991, 1995). However, the issue of whether to keep ratings is more complex and goes beyond issues of psychometric quality and utility.

Proponents of ratings have articulated valid concerns about the need to document performance to meet the legal and regulatory requirements that exist in some countries. They have argued that evaluation occurs irrespective of whether ratings are recorded and that there is value in the transparency ratings provide. Opponents alternatively argue that it is worse to provide ratings that are almost always inflated but at the same time maintain nontransparent sources of more accurate performance information. For example, designations of who is high potential are not often shared with employees. The most compelling rationale supporting ratings and PM processes in general is that while these are ineffective in driving performance, they at least help ensure that some performance information is communicated to employees (Adler et al. 2016). Supporting this notion is recent CEB (2016) research, which shows that employees in rating-less organizations report less engagement and perceived fairness than employees who received ratings. However, a cautionary note in interpreting these results is that CEB's research did not separately evaluate organizations that replaced ratings with regular check-ins, coaching, and feedback from those that simply eliminated ratings. Thus, although CEB's findings suggest negative attitudes may be associated with no ratings, these may be more a function of poor practices that remained unaddressed when ratings were removed. It is not recommended that organizations remove ratings

DESIGNING PERFORMANCE MANAGEMENT PROCESSES TO DRIVE PERFORMANCE

Performance management (PM) practices should be designed to drive performance in support of the organization's overall strategy and goals. Taking a strategic approach to PM design requires answering questions like the following:

- What business problem(s) need to be solved?
- What must be evaluated to achieve our business goals?
- How much pay is at risk?
- Will a common process work, or do we need different processes for different work and units?

without adding other mechanisms or processes to ensure employees receive clear expectations and regular feedback.

Although the question of whether to retain ratings has spawned great interest and lively debate, this is not the most important or first question to ask in considering PM transformation (Adler et al. 2016). More important questions concern the critical performance outcomes an organization wants to drive and how employees can be best enabled to deliver these. From this view, there is no right answer to the question of whether an organization should have ratings. The answer depends on the organization's strategy and goals, reward structures, maturity, openness to change, and management philosophy, among other contextual factors (see sidebar Designing Performance Management Processes to Drive Performance). Pulakos et al. (2015) argued for taking a strategic approach to PM in which PM goals emanate from strategic goals and then choices about system design are made based on these criteria. Taking a more strategic view versus a narrow view that overemphasizes the question of ratings will help organizations think through how to best drive performance in their given context. To the extent that organizations have different job levels, work requirements, and outcomes across different units, different PM features or processes may be needed to optimize performance versus a one-size-fits-all model (Church et al. 2015).

Much of the evidence to date on the efficacy of PM transformation efforts is based on individual case studies. However, the Center for Effective Organizations (CEO) recently conducted research that begins to provide insights on the effectiveness of different innovative practices. Three new practices, rating-less reviews, ongoing feedback, and crowdsourced feedback that leverages social media, have been adopted by enough organizations in the past several years that CEO was able to begin evaluating their impact (Ledford et al. 2016). On the basis of survey feedback from 244 companies, CEO found that almost all (97%) had adopted ongoing feedback, fewer (51%) had adopted rating-less reviews, and the least number of companies (27%) had adopted crowdsourced feedback. Interestingly, these new practices did not necessarily replace legacy practices, as many companies continued use of cascading goals, 360-degree feedback, competency assessment, and rating calibration, along with at least one of the cutting-edge practices. However, the use of ongoing feedback and rating-less reviews in combination was associated with decreased use of more traditional PM practices.

The results of CEO's research showed that the combination of all three innovative practices yielded the most impact on several criteria of interest, including strategic alignment, motivating and developing employees, and rewarding top talent. The combination of ongoing feedback and crowdsourced feedback was more impactful than either ongoing feedback alone or ongoing feedback plus rating-less reviews. Responses to the survey of innovative practices compared to the responses from prior survey research on PM practices suggested that the new practices are more

effective than traditional practices. These results are encouraging and suggest promise for the direction companies are taking to improve their PM practices. The important question is whether the increased effectiveness associated with the new practices will sustain over time.

Driving More Effective Performance Management Behavior

The second strategy organizations are using to improve the value of PM is through changing manager and employee behavior. The importance of manager behavior and employee-manager interpersonal relationships on employee attitudes and performance has been widely recognized and is supported by decades of research. For example, studies have shown that effective relationships and open communication are associated with perceptions of fairness and procedural justice (Beer 1981, DeCotiis & Petit 1978, Wexley & Klimoski 1984) as well as higher performance (Daniels 2000, Beer 1981). Best practice survey research has similarly shown the importance of effective communication and solid manager-employee relationships as levers for engagement and high performance (CEB 2004, Harter et al. 2002).

The importance of feedback, in particular, has consistently been highlighted in the research literature (Bernardin & Beatty 1984, Ilgen et al. 1979, Maier 1958), because of its potential to positively impact performance (Ilgen et al. 1979, Kluger & DeNisi 1996) and job attitudes (Ilgen et al. 1981, Pearson 1991). Although researchers have invested considerable energy in determining how to most effectively conduct formal performance feedback sessions (e.g., Burke et al. 1978, Cederblom 1982, Pearce & Porter 1986, Nathan et al. 1991), it is now widely understood that informal feedback provided in real time is most impactful in driving performance and engagement (e.g., Aguinis 2013, CEB 2004, Kirkland & Manoogian 2007, Gregory et al. 2008).

Effective PM behavior goes beyond informal feedback, however (see sidebar Key Performance Management Behaviors for Managers). CEB (2004) showed that several key manager behaviors are associated with higher performance and engagement: setting clear expectations, providing regular informal feedback, and helping employees develop and succeed at work. Google's Project Oxygen (Bryant 2011) identified eight habits of highly effective managers that confirmed CEB's findings. These included meeting individually with employees, helping them solve problems, and coaching and developing them. Google managers who demonstrated these behaviors had teams that performed better, stayed longer, and had better attitudes about work. Google began teaching the eight habits in training programs, as well as in coaching and performance review sessions with individual managers. These efforts yielded significant improvement in 75% of Google's least effective managers. Taken together, these studies demonstrate the strong effects of manager behavior on employee performance, engagement, and bottom-line results, and they support the idea that

KEY PERFORMANCE MANAGEMENT BEHAVIORS FOR MANAGERS

Below are the key behaviors research has shown that managers need to exhibit to drive high employee performance and engagement.

- Set clear expectations, priorities, success criteria, and standards.
- Revise expectations in real time, so employees know what to do.
- Provide informal feedback daily to praise, coach, and course-correct employee performance.
- Check in regularly with employees to stay in touch and provide guidance.
- Coach employees and help them solve problems to enable success.

KEY PERFORMANCE MANAGEMENT BEHAVIORS FOR EMPLOYEES

Below are the key behaviors employees need to exhibit to do their part in driving high performance.

- Clarify their performance expectations to ensure they understand priorities and standards; revisit expectations when necessary.
- Set expectations with peers about who is doing what, and by when.
- Ask for and accept feedback openly and nondefensively.
- Use feedback to course-correct and continuously improve own performance.

key behaviors can be learned with proper training and coaching. When the significant impact of manager behavior is compared to the low impact formal PM systems have on performance (CEB 2004, Bryant 2011, Pulakos et al. 2015), it is not surprising that PM reform today focuses heavily on behavior change.

Although research has primarily focused on critical manager behaviors, it is important to remember that PM is an interactive process between managers and employees in which employees also have responsibilities for enabling performance. The aspiration of PM behavior change is to fundamentally shift how PM is viewed and carried out. This requires managers and employees to make both mindset and behavioral changes so that PM is transformed from a formal HR system of prescribed steps that are cued by an automated system to managers and employees engaging together on a continuous basis to drive high performance and achieve important business outcomes (see sidebar Key Performance Management Behaviors for Employees). **Figure 2** illustrates the desired change, which is rooted in developing productive working relationships that are characterized by open communication and trust, which in turn enable openness to real-time feedback, coaching, continuous learning, and development to occur naturally as part of daily work. In essence, a new mindset and climate need to be created about how PM is enacted on a regular cadence through key manager and employee behaviors (Bryant 2011, CEB 2004, Pulakos et al. 2015).

Creating a climate for PM that is characterized by effective PM behavior is no easy task. Many approaches oversimplify what is required, which risks yet another set of failed PM interventions (Mueller-Hanson & Pulakos 2018). For example, some companies have attempted to drive informal feedback by simply requiring more frequent scheduled check-ins between managers and employees. Others have implemented formal training programs but have done too little, or virtually nothing, to enable the deep learning and change that need to transfer into the work environment.

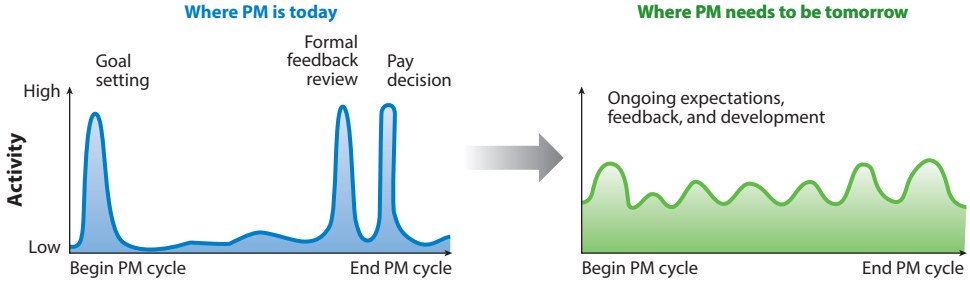

Figure 2

Performance management transformation to improve its effectiveness and value.

Behavior change needs to go beyond passive methods, such as e-learning, classroom learning, and even more active methods such as simulated practice (e.g., role plays). Meaningful change requires viewing the concept of PM differently, engaging differently, and reacting differently over time and across contexts until new behavioral patterns are formed and embedded.

Experiential learning on the job provides a platform for behavior change (DeRue et al. 2012, Ericsson et al. 1993), because work inherently contains several important drivers of deep learning (Davachi et al. 2010), as follows:

- Work tasks inherently capture learners' attention and have built-in relevance and ownership, which are important to hardwire learning.
- Practicing new concepts as part of real work helps learners see how these play out across people and situations, providing contextualization, personalization, and varied learning experiences that are essential for deep learning.
- Work provides natural opportunities for spaced practice that solidifies learning.
- Learning through real work incorporates important social elements, such as learning generation by participants, connectedness, and opportunities for feedback.

Pulakos et al. (2015) offered an experiential on-the-job learning approach to PM behavior change that prompts learners to practice target behaviors, assists them in extracting learning through reflection, and reinforces repetition to hardwire learning over time until new behaviors are performed successfully and nearly automatically, that is, without significant cognitive or emotional strain (Fitts & Posner 1967). This approach was shown to result in positive behavior change as judged by those with whom trainees interacted on the job (Mueller-Hanson et al. 2016). However, in some situations in which this intervention was attempted, there was insufficient attention and staying power to drive meaningful change.

Future research is needed to continue evaluating different approaches to large-scale PM behavior change. As one example, Boyatzis et al. (2015) offered a comprehensive model of change that may better enable sustained behavior change in organizations. To be successful, individuals must genuinely care about making a change, engage in self-directed, intentional effort to make a change, and be aware of the change and the process of change. In the model, change begins by creating a personal future vision that provides the impetus for change. The personal vision represents the ideal self, which is then compared to the real self (which is partly a function of how one is viewed by others) to identify strengths, which is where the ideal and real selves are consistent, and gaps. The gap between the real and the ideal selves creates the agenda for learning and provides motivation to change. Successful change is more likely to occur when individuals adopt a learning versus performance orientation to arouse positive belief in one's capability to change and maintain focus on what the individual wants to become. By contrast, a performance orientation that focuses on success, proving oneself, and getting praise arouses anxiety in the wrong parts of the brain to promote learning and development. The idea is that individuals must become intrinsically motivated to change, driven by the emotional part of their brains, versus feeling obligated to change because others told them to change. Also important are trusting and supportive relationships to enable the change process. Boyatzis (2016) cites several studies in which this approach yielded more successful behavior change outcomes than use of more traditional learning approaches.

Effective PM behavior may also be facilitated by mindfulness practices that are gaining increasing attention in organizations to enable higher performance, satisfaction, and resilience, among other positive work outcomes (Gelles 2012, Dane 2011, Dane & Brummel 2013). The key tenants of mindfulness include the practice of being present, being aware of what is happening in the moment, pausing to control fight-or-flight reactions, and using this awareness to make better decisions. Mueller-Hanson & Pulakos (2018) have hypothesized that these practices should

help managers be more in tune with the extent to which they are setting meaningful expectations that are well-understood by employees and help them naturally pause and check on how expectations and feedback are being received. They may also help increase managers' awareness of the impact they are having on employees and develop more effective coaching skills. Mindfulness practices should help employees become more attuned to feedback cues and take in feedback more openly by avoiding automatic defensive reactions, separating the message from their reaction to it, and processing information more effectively to enhance learning. One potential challenge is that developing mindfulness requires engaging in a set of practices over time that enable it. Future research is needed in organizational settings to determine whether these practices can in fact drive more effective PM behavior and, if so, how they can best be implemented to yield the intended results.

There are three cautionary notes about successful behavior change to bear in mind. The first is that many companies have unrealistic expectations about how long and what it takes for meaningful change to occur, which includes intentionality from learners, a learning orientation, and repeated practice (Pulakos et al. 2015). The second caution is that PM behavior change (e.g., real-time feedback, agile goal setting, coaching, etc.) is unlikely to be successfully embedded without context. This is because concepts such as agile goal setting, real-time feedback, and effective coaching are elusive ideas that cannot be well understood and do not gain sufficient meaning to be learned until they are practiced in the actual work context. The third caution, which is related to the first two, is that new PM behaviors will be most successfully embedded by intentionally applying them in support of achieving an important performance goal (Davachi et al. 2010, Mueller-Hanson & Pulakos 2018), because this enables them to be learned as they will be used—as enablers of high performance. A case study from Alcoa in which new, keystone habits were successfully developed to improve safety illustrates the importance of attaching learning to concrete, well-understood outcomes that have high relevance to individuals and the business (Duhigg 2012).

PM behavior change should thus start with identification of a concrete performance goal (e.g., improve collaboration within the team, serve customers faster). Employees and managers would need to have honest conversations about what is getting in the way of achieving optimal performance today, and real-time feedback would need to be given when instances of poor and effective performance occur. Although individual manager-employee pairs can select a performance goal on which to practice and learn effective PM behaviors, the impact of behavior change interventions will be most powerful when whole teams or organizations develop a meaningful shared vision and engage in learning and behavior change together, as this increases the likelihood of developing the momentum and sustainability that are needed for real change to occur (Boyatzis 2016). Combining behavior change interventions with other strategies discussed above may be important to achieve maximum performance impacts, for example, incorporating natural feedback methods (e.g., crowdsourcing), leveraging relevant performance metrics that are available in the environment, and focusing on team rather than individual outcomes and performance criteria where this makes sense.

Successful Implementation of Performance Management Processes

Although we touched on the topic of implementation above, PM transformation is a major organizational change effort, and established change management models provide insights about how to orchestrate change effectively (Kotter 2007, Bridges & Bridges 2016, Heath & Heath 2010, Cohen 2005). As with any change effort, success requires developing a business case, engaging stakeholders, gaining buy-in and support, communicating effectively, enabling and empowering the workforce to embed the change, and managing expectations. An important aspect of

implementation, however, is ensuring that the PM approach fits the organization's strategy and culture, and that it can be successfully implemented within the given context.

More so than any other talent management system, PM is characterized by almost blind adoption of new trends that frequently underdeliver on expectations (Pulakos & O'Leary 2011). This happens because insufficient consideration is given to how well a new practice actually fits with the organization's strategy, culture, appetite, and resources. An example provided by Mueller-Hanson & Pulakos (2018) illustrates this point. In *Work Rules*, Bock (2015) describes how Google's practice of rating and paying high and low performers very differently fits with Google's culture—which is highly data-driven, offers big rewards, and has tolerance for lengthier processes that are required to make more nuanced distinctions in pay (Bock 2015). However, in other organizations, pay-for-performance systems have been abysmal failures, when performance has not been differentiated historically, the culture is more egalitarian, and insufficient variable pay exists to make meaningful differentiated rewards. The point is simply that every aspect of a PM process needs to fit well within the specific organizational context in which it is implemented to be successful.

PM change that requires thinking and behaving differently takes more time than changing the mechanics of a process. The former can take years, whereas the latter can often be done in weeks. Behavior and culture change that builds incrementally over time—starting small, proving the concept, and using success to build momentum—tends to be more successful in achieving long-term change (Mueller-Hanson & Pulakos 2018). Employees need to be actively engaged in creating a vision for their personal change, which is essential for building the motivation to try out and use new behaviors (Boyatzis et al. 2015). Repetition, reinforcement, and patience are important in attitude, behavior, and culture change. It is important to be realistic about what outcomes can be achieved in a given context, as well as context factors that shape what should be attempted and can be accomplished. This means realistically assessing political, social, and motivational factors that will enable or undermine change, as these are often given insufficient attention. Although some contextual factors can be effectively mitigated, others may be intractable barriers to success. Very significant barriers to success need to be addressed before change is attempted.

Summary and Next Steps for Performance Management Research and Practice

Over almost a century, PM has evolved from a narrow focus on performance evaluation to the design of comprehensive annual PM processes. What was initially conceived of as a relatively simple problem of how to define rating criteria that would enable managers to make accurate ratings has turned into a much more complex challenge of understanding and studying a vast array of multifaceted political, social, motivational, environmental, and practical factors that drive PM behavior and outcomes. This metamorphosis of research and practice over time raises questions about whether PM processes can add value in organizations or will continue to be plagued by intractable challenges. Most of the many attempts to improve PM outcomes and attitudes have focused on implementing formal steps, tools, and processes to align goals, improve rating accuracy, address performance gaps, and make better decisions about employees. Unfortunately, this focus on the formal PM system has yielded largely disappointing results in terms of driving important performance outcomes—which should be the ultimate purpose of PM processes.

Thought leaders have successfully argued that attention needs to be directed to informal processes and behaviors that facilitate performance day to day rather than overengineering formal systems that sit outside daily work. Specific behaviors that have been suggested as most important include setting agile goals so that expectations can be adjusted as needed to remain clear and relevant, providing real-time feedback that contributes to learning and achieving more effective results, and removing barriers to help employees accomplish their goals. We acknowledge that

behavior change is difficult—on par with the difficulty and complexity of any major organizational change that requires substantial resource and time investments to achieve. Thus far, effective, sustainable behavior change in organizations has largely eluded us, although several behavior change strategies from the clinical, medical, and neuroscience domains appear to hold promise for driving more effective PM. Although wholesale behavior and culture change to drive performance may be the aspiration, recall that positive impacts can be achieved with lesser interventions. As Ledford et al. (2016) have shown, even one or two well-considered changes to streamline the formal system or add new features such as rating-less reviews and crowdsourced feedback can add value and result in positive outcomes.

An important area for future research to assess is the extent to which interventions to streamline formal PM systems and drive more effective PM behavior will yield sustained improvements in managing performance. This is important because, historically, PM has been susceptible to implementation of new trends that yield initial benefit but fizzle over time. Questions remain about whether the latest efforts to reduce process and drive effective behavior will finally move the needle in delivering more effective PM practices or will simply end up as the next PM trend that ultimately falls short. Several authors, but most notably Murphy & Cleveland (1995), have convincingly argued that strong political and social drivers of manager behavior are often at odds with what managers are asked to do in formal PM processes. These systems can thus be counterproductive for managers and interfere with their ability to get work done, which adds further insight into why PM has likely been so challenged in producing value.

Although these conflicts have been well articulated in the literature, relatively little work has been directed at addressing them head-on. Toward this end, it may be useful to consider new models that focus on playing to employee strengths and accepting that everyone has limitations within bounds rather than imposing requirements to provide developmental feedback and improve performance areas that may not be easily addressed. Similarly, we know that organizations are becoming flatter and more work is being executed collaboratively through matrixed, cross-functional teams (CEB 2012). Although the importance of teams in executing work has been discussed extensively, relatively few organizations have implemented team-based PM processes. However, it may be more productive for managers to start thinking in terms of the collection of skills and characteristics they need to execute their work programs and strategies for managing performance that are more team-based. These ideas are not to suggest that individual performance be forgotten or poor performance be tolerated, but there may be room to shift our thinking to models that would better support the human and practical realities of the imperfect performers in the workplace that collectively need to get a job done. Pink (2009) and Rock (2008) have both offered several ideas that may hold promise for reducing conflict between PM requirements and how managers and employees naturally engage in work.

A final area that seems promising for future work is more rigorous definition of the context in which PM occurs and identification of which context factors matter most for effective PM process and practice. Beyond political, social, and practical factors that impact ratings, Church et al. (2015) have gone a step further in suggesting that different PM processes may be required to account for the different contexts that exist within an organization. For example, differences in work requirements, complexity, goals, and levels, among many other factors, may have important implications for PM design. Designers of talent management systems are beginning to use context factors directly to fine-tune talent management practices and decisions. In the leadership area, for example, defining the context in which leaders need to perform enables matching leaders to different roles with substantially higher predictive accuracy than when context factors are not considered (CEB 2017). Thus, a final area for future research is to define and test the impact of a fuller set of contextual factors on PM effectiveness and potentially leverage these insights directly to more

SUMMARY: PERFORMANCE MANAGEMENT

Below is a summary of our current conclusions regarding performance management (PM) based on research and practice to date:

- Formal PM processes disengage employees, cost millions, and have no impact on performance.
- Formal systems can be streamlined but should not be eliminated without robust informal processes.
- Informal day-to-day PM behaviors enable performance but take time and effort to embed.
- Future research should investigate new PM models that leverage neuroscience, are strength-based, and focus on team performance.
- The impact of context factors on the design and effectiveness of PM should be further evaluated.

systematically design future processes that are best fit to the contextual factors that are most important in a given situation. Key conclusions regarding the current status of performance management research and practice are summarized (see sidebar Summary: Performance Management).

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Adler S, Campion M, Colquitt A, Grubb A, Murphy K, et al. 2016. Getting rid of performance ratings: Genius or folly? A debate. *Ind. Organ. Psychol.* 9(2):219–52

Aguinis H. 2013. *Performance Management*. Upper Saddle River, NJ: Pearson/Prentice Hall. 3rd ed.

Baldassarre L, Finken B. 2015. GE's real-time performance development. *Harvard Business Review*, Aug. 12. **https://hbr.org/2015/08/ges-real-time-performance-development**

Bartram D. 2007. Increasing validity with forced-choice criterion measurement formats. *Int. J. Sel. Assess.* 15:263–72

Beer M. 1981. Performance appraisal: dilemmas and possibilities. *Organ. Dyn.* 9(3):24–36

Bernardin HJ, Beatty RW. 1984. *Performance Appraisal: Assessing Human Behavior at Work*. Boston: Kent

Blanz F, Ghiselli EE. 1972. The mixed standard scale: a new rating system. *Pers. Psychol.* 25:185–99

Bock L. 2015. *Work Rules! Insights from Google that Will Transform How You Live and Lead*. New York: Twelve

Borman WC. 1974. The rating of individuals in organizations: an alternate approach. *Organ. Behav. Hum. Perform.* 12:105–24

Borman WC. 1975. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *J. Appl. Psychol.* 60(5):556–60

Borman WC. 1977. Consistency of rating accuracy and rating errors in the judgment of human performance. *Organ. Behav. Hum. Perform.* 20:238–52

Borman WC, Buck E, Hanson MA, Motowidlo SJ, Stark S, Drasgow F. 2001. An examination of the comparative reliability, validity, and accuracy of performance ratings made using computerized adaptive rating scales. *J. Appl. Psychol.* 86:965–73

Boyatzis RE. 1996. Consequences and rejuvenation of competency-based human resource and organization development. In *Research in Organizational Change and Development*, Vol. 9, ed. RW Woodman, WA Pasmore, pp. 101–22. Greenwich, CN: JAI Press

Boyatzis RE. 2016. *Unleashing the Power of Intentional Change*. New York: Korn Ferry, Hay Group

Boyatzis RE, Rochford K, Taylor SN. 2015. The role of the positive emotional attractor in vision and shared vision: toward effective leadership, relationships, and engagement. *Front. Psychol.* 6:670

Bracken D, Timmreck C, Church A, eds. 2001. *Handbook of Multisource Feedback*. San Francisco: Jossey-Bass
Bridges W, Bridges S. 2016. *Managing Transitions: Making the Most of Change*. Boston: Perseus Book Group
Bryant A. 2011. Google's quest to build a better boss. *New York Times*, Mar. 12. **https://www.nytimes.com/2011/03/13/business/13hire.html**
Buckingham M, Goodall A. 2015. Reinventing performance management. *Harvard Business Review*, April. **https://hbr.org/2015/04/reinventing-performance-management**
Burke RJ, Weitzel W, Weir T. 1978. Characteristics of effective employee performance review and development interviews: replication and extension. *Pers. Psychol.* 31(4):903–19
Cascio WF. 1998. *Applied Psychology in Human Resource Management*. Upper Saddle River, NJ: Prentice Hall
CEB. 2004. *Driving employee performance and retention through engagement: a quantitative analysis of the effectiveness of employee engagement strategies*. Washington, DC: CEB. **https://www.stcloudstate.edu/humanresources/_files/documents/supv-brown-bag/employee-engagement.pdf**
CEB. 2012. *Driving breakthrough performance in the new work environment*. Washington, DC: CEB. **http://blueroom.neuroleadership.com/assets/documents/CEB%20-%20Driving%20Breakthrough%20Performance%20in%20the%20New%20Work%20Environment%20-%20Short%20Version.pdf**
CEB. 2016. *The real impact of eliminating performance ratings: insights from employees and managers*. Washington, DC: CEB. **https://www.cebglobal.com/content/dam/cebglobal/us/EN/best-practices-decision-support/human-resources/pdfs/eliminate-performance-ratings.pdf**
CEB. 2017. *The power of context in driving leader success*. Washington, DC: CEB. **https://www.cebglobal.com/corporate-leadership-council/power-of-context-in-driving-leader-success.html**
Cederblom D. 1982. The performance appraisal interview: a review, implications, and suggestions. *Acad. Manag. Rev.* 7(2):219–27
Church AH, Ginther NM, Levine R, Rotolo CT. 2015. Going beyond the fix: taking performance management to the next level. *Ind. Organ. Psychol.* 8:121–29
Cohen D. 2005. *The Heart of Change Field Guide: Tools and Tactics for Leading Change in Your Organization*. Boston: Harv. Bus. Rev. Press
Cooper WH. 1981. Ubiquitous halo. *Psychol. Bull.* 90:218–44
Culbert SA, Rout L. 2010. *Get Rid of the Performance Review: How Companies Can Stop Intimidating, Start Managing—And Focus on What Really Matters*. New York: Bus. Plus
Culbertson SS, Henning JB, Payne SC. 2013. Performance appraisal satisfaction: the role of feedback and goal orientation. *J. Pers. Psychol.* 12:189–95
Cunningham L. 2015. In big move, Accenture will get rid of annual performance reviews and rankings. *Washington Post*, July 21. **http://www.washingtonpost.com/blogs/on-leadership/wp/2015/07/21/in-big-move-accenture-will-get-rid-of-annual-performance-reviews-and-rankings/?tid=pm_pop_b**
Dane E. 2011. Paying attention to mindfulness and its effects on task performance in the workplace. *J. Manag.* 37:997–1018
Dane E, Brummel BJ. 2013. Examining workplace mindfulness and its relations to job performance and turnover intention. *Hum. Relat.* 67:105–28
Daniels AC. 2000. *Bringing Out the Best in People: How to Apply the Astonishing Power of Positive Reinforcement.* New York: McGraw-Hill
Davachi L, Kiefer T, Rock D, Rock L. 2010. Learning that lasts through AGES. *NeuroLeadership J.* 3:53–63
DeCotiis T, Petit A. 1978. The performance appraisal process: a model and some testable propositions. *Acad. Manag. Rev.* 3:635–46
DeNisi AS. 2006. *A Cognitive Approach to Performance Appraisal*. New York: Routledge
DeNisi AS, Murphy KR. 2017. Performance appraisal and performance management: 100 years of progress? *J. Appl. Psychol.* 3:421–33
DeRue DS, Nahgang JD, Hollenbeck JR, Workman K. 2012. A quasi-experimental study of after-event reviews and leadership development. *J. Appl. Psychol.* 97:997–1015
Dubois DD. 1993. *Competency-Based Performance Improvement: A Strategy for Organizational Change*. Amherst, MA: HRD Press
Duhigg C. 2012. *The Power of Habit: Why We Do What We Do*. New York: Random House

Dunnette MD. 1963. A note on the criterion. *J. Appl. Psychol.* 47:251–54
Drucker PF. 1954. *The Practice of Management*. New York: Harper & Row
Effron M, Ott M. 2010. *One Page Talent Management: Eliminating Complexity, Adding Value*. Cambridge, MA: Harv. Bus. Press
Ericsson KA, Krampe RT, Tesch-Romer C. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100:363–406
Feldman JM. 1981. Beyond attribution theory: cognitive processes in performance appraisal. *J. Appl. Psychol.* 66(2):127–48
Fitts PM, Posner MI. 1967. *Human Performance*. Oxford, UK: Brooks/Cole
Gelles D. 2012. *Mindful Work: How Meditation Is Changing Business from the Inside Out*. New York: Houghton Mifflin Harcourt
Gregory JB, Levy PE, Jeffers M. 2008. Development of a model of the feedback process within executive coaching. *Consult. Psychol. J.* 60(1):42–56
Greguras GJ, Robie C, Schleicher DJ, Goff M. 2003. A field study of the effects of rating purpose on the quality of multisource ratings. *Pers. Psychol.* 56(1):1–21
Grote RC. 1996. *The Complete Guide to Performance Appraisal*. New York: Am. Manag. Assoc.
Guion RM. 1961. Criterion measurement and personnel judgments. *Pers. Psychol.* 4:141–49
Harter JK, Schmidt FL, Hayes TL. 2002. Business unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. *J. Appl. Psychol.* 87(2):268–79
Heath D, Heath C. 2010. *Switch: How to Change Things When Change Is Hard*. New York: Broadway Books
Hillgren JS, Cheatham DW. 2000. *Understanding Performance Measures: An Approach to Linking Rewards to the Achievement of Organizational Objectives*. Scottsdale, AZ: WorldatWork
Ilgen DR, Feldman JM. 1983. Performance appraisal: a process focus. In *Research in Organizational Behavior*, Vol. 5, ed. L Cummings, B Staw, pp. 141–97. Greenwich, CT: JAI Press
Ilgen DR, Fisher CD, Taylor MS. 1979. Consequences of individual feedback on behavior in organizations. *J. Appl. Psychol.* 64(4):349–71
Ilgen DR, Peterson RB, Martin BA, Boeschen DA. 1981. Supervisor and subordinate reactions to performance appraisal sessions. *Organ. Behav. Hum. Perform.* 28(3):311–30
Institute for Corporate Productivity (i4cp). 2011. Tying pay to performance. *i4cp*, Sept. 8. **https://www.i4cp.com/productivity-blog/2011/09/08/companies-tying-employee-pay-to-performance-increases-17-in-last-two-years**
Jamieson BD. 1973. Behavioral problems with management by objective. *Acad. Manag. Rev.* 16:496–505
Kantor J, Streitfeld D. 2015. Inside Amazon: wrestling big ideas in a bruising workplace. *New York Times*, Aug. 15. **https://www.nytimes.com/2015/08/16/technology/inside-amazon-wrestling-big-ideas-in-a-bruising-workplace.html**
Kirkland K, Manoogian S. 2007. *Ongoing feedback, how to get it, how to use it*. Greensboro, NC: Cent. Creat. Leadersh.
Klein AL. 1996. Validity and reliability for competency-based systems: reducing litigation risks. *Compens. Benefits Rev.* 28(4):31–37
Kluger AN, DeNisi AS. 1996. The effects of feedback interventions on performance: a historical review, meta-analysis and a preliminary feedback intervention theory. *Psychol. Bull.* 119:254–84
Kotter JP. 2007. Leading change: why transformation efforts fail. *Harvard Business Review*, Jan. **https://hbr.org/2007/01/leading-change-why-transformation-efforts-fail**
Landy FJ, Farr JL. 1980. Performance rating. *Psychol. Bull.* 87:72–107
Landy FJ, Farr JL. 1983. *The Measurement of Work Performance*. New York: Academic
Latham G, Wexley K. 1977. Behavioral observation scales. *J. Appl. Psychol.* 30:255–68
Latham GP, Locke EA. 2007. New developments and directions for goal-setting research. *Eur. Psychologist* 123:290–300
Latham GP, Wexley KN, Pursell ED. 1975. Training managers to minimize rating errors in the observation of behavior. *J. Appl. Psychol.* 60(5):550–55
Ledford GE, Benson GS, Lawler EE III. 2016. *A Study of Cutting-Edge Performance Management Practices: Ongoing Feedback, Ratingless Reviews and Crowdsourced Feedback*. Scottsdale, AZ: WorldatWork

Locke EA, Latham GP. 1990. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice-Hall

Locke EA, Shaw KN, Saari LM, Latham GP. 1981. Goal setting and task performance, 1969–1980. *Psychol. Bull.* 90:125–52

London M, Mone EM. 2014. Performance management processes that reflect and shape organizational culture and climate. In *The Oxford Handbook of Organizational Climate and Culture*, ed. B Schneider, KM Barbera, pp. 79–100. Oxford, UK: Oxford Univ. Press

Lopez FM. 1968. *Evaluating Employee Performance*. Chicago, IL: Public Personnel Association

Maier NR. 1958. *The Appraisal Interview: Objectives, Methods and Skills*. New York: Wiley

McIntyre RM, Smith DE, Hassett CE. 1984. Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *J. Appl. Psychol.* 69(1):147–56

Mosley E. 2015. Creating an effective peer review system. *Harvard Business Review*, Aug. 19. **https://hbr.org/2015/08/creating-an-effective-peer-review-system?cm_sp=Article-_-Links-_-Top%20of%20Page%20Recirculation**

Mueller-Hanson RA, Garza M, Riordan BG. 2016. *Performance Management Transformation*. Arlington, VA: CEB

Mueller-Hanson RA, Pulakos ED. 2018. *Transforming Performance Management to Drive Performance: An Evidence-based Roadmap*. New York: Routledge

Murphy KR, Balzer WK. 1989. Rater errors and rating accuracy. *J. Appl. Psychol.* 74:619–24

Murphy KR, Cleveland JN. 1991. *Performance Appraisal. An Organizational Perspective*. Needham Heights, MA: Allyn and Bacon

Murphy KR, Cleveland JN. 1995. *Understanding Performance Appraisal: Social, Organizational and Goal-Oriented Perspectives*. Newbury Park, CA: Sage

Murphy KR, Cleveland JN, Skattebo AL, Kinney TB. 2004. Raters who pursue different goals give different ratings. *J. Appl. Psychol.* 89:158–64

Murphy KR, Jako RA, Anhalt RL. 1993. Nature and consequences of halo error: a critical analysis. *J. Appl. Psychol.* 78:218–25

Murphy KR, Martin C, Garcia M. 1982. Do behavioral observation scales measure observation? *J. Appl. Psychol.* 67:562–67

Nathan BD, Mohrman AM, Milliman J. 1991. Interpersonal relations as a context for the effects of appraisal interviews on performance and satisfaction: a longitudinal study. *Acad. Manag. J.* 34(2):352–69

O'Boyle E, Aguinis H. 2012. The best and the rest: revisiting the norm of normality of individual performance. *Pers. Psychol.* 65:79–119

O'Leary RS, Pulakos ED. 2017. Defining and measuring results of workplace behavior. In *The Handbook of Employee Selection*, ed. JL Farr, N Tippins, pp. 509–29. New York: Psychology Press. 2nd ed.

Patterson DG. 1922. The Scott Company graphic rating scale. *J. Pers. Res.* 1:361–76

Pearce JL, Porter LW. 1986. Employee responses to formal performance appraisal feedback. *J. Appl. Psychol.* 71(2):211–18

Pearson CAL. 1991. An assessment of extrinsic feedback on participation, role perceptions, motivation, and job satisfaction in a self-managed system for monitoring group achievement. *Hum. Relat.* 44(5):517–37

Pink DH. 2009. *Drive: The Surprising Truth About What Motivates Us*. New York: Riverhead Books

Ployhart RE, Weekley JA, Ramsey J. 2009. The consequences of human resource stocks and flows: a longitudinal examination of unit service orientation and unit effectiveness. *Acad. Manag. J.* 52(5):996–1015

Pulakos ED. 1984. A comparison of rater training programs: error training and accuracy training. *J. Appl. Psychol.* 69:581–88

Pulakos ED. 1986. The development of a training program to increase accuracy with different rating formats. *Organ. Behav. Hum. Decis. Process.* 38:76–91

Pulakos ED. 2009. *Performance Management: A New Approach for Driving Business Results*. Oxford: Wiley-Blackwell

Pulakos ED, Mueller-Hanson RA, Arad S, Moye N. 2015. Performance management can be fixed: an on-the-job experiential learning approach for complex behavior change. *Ind. Organ. Psychol.* 8:51–76

Pulakos ED, Mueller-Hanson RA, O'Leary RS, Meyrowitz MM. 2012. *Practice Guidelines: Building a High-Performance Culture: A Fresh Look at Performance Management*. Alexandria, VA: Soc. Hum. Resour. Manag. Found.

Pulakos ED, O'Leary RS. 2010. Defining and measuring results of workplace behavior. In *The Handbook of Employee Selection*, ed. JL Farr, N Tippins, pp. 513–29. New York: Psychology Press

Pulakos ED, O'Leary RS. 2011. Why is performance management so broken? *Ind. Organ. Psychol.* 4(2):146–64

Rock D. 2008. SCARF: a brain-based model for collaborating with and influencing others. *NeuroLeadership J.* 1:44–52

Rock D, Jones B. 2015. Why more and more companies are ditching performance ratings. *Harvard Business Review*, Sept. 8. **https://hbr.org/2015/09/why-more-and-more-companies-are-ditching-performance-ratings**

Rodgers R, Hunter JE. 1991. Impact of management by objectives on organizational productivity. *J. Appl. Psychol.* 76:322–36

Rodgers R, Hunter JE, Rogers DL. 1993. Influence of top management commitment on management process success. *J. Appl. Psychol.* 78:151–55

Saal FE, Landy FJ. 1977. The mixed standard rating scale: an evaluation. *Organ. Behav. Hum. Perform.* 18:19–35

Schippmann JS. 1999. *Strategic Job Modeling: Working at the Core of Integrated Human Resource Systems*. Mahwah, NJ: Lawrence Erlbaum Assoc.

Schneider RJ, Goff M, Anderson S, Borman WC. 2003. Computerized adaptive rating scales for measuring managerial performance. *Int. J. Sel. Assess.* 11:237–46

Schneier CE, Shaw DG, Beatty RW. 1991. Performance measurement and management: a tool for strategy execution. *Hum. Resour. Manag.* 30:279–301

Scott WD, Clothier RC, Spriegel WR. 1941. *Personnel Management*. New York: McGraw-Hill

Smith PC, Kendall LM. 1963. Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. *J. Appl. Psychol.* 47:149–55

Smither JW, London M. 2009. *Performance Management: Putting Research into Action*. New York: Wiley

Smither JW, London M, Reilly RR. 2005. Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Pers. Psychol.* 58:33–66

Strauss G. 1972. Management by objectives: a critical review. *Train. Dev. J.* 26:10–15

Taylor FW. 1911. *The Principles of Scientific Management*. New York: Harper & Brothers

Tziner A, Murphy K. 1999. Additional evidence of attitudinal influences in performance appraisal. *J. Bus. Psychol.* 13:407–19

Viswesvaran C, Ones DS, Schmidt FL. 1996. Comparative analysis of the reliability of job performance ratings. *J. Appl. Psychol.* 81:557–74

Wexley KN, Klimoski R. 1984. Performance appraisal: an update. In *Research in Personnel and Human Resources Management*, ed. KM Rowland, GD Ferris, pp. 35–79. Greenwich, CN: JAI Press

Winters D, Latham GP. 2006. The effect of learning versus outcome goals on a simple versus a complex task. *Group Organ. Manag.* 21:236–50