Focal Article

Getting Rid of Performance Ratings: Genius or Folly? A Debate

Seymour Adler
Aon Hewitt

Michael Campion
Purdue University

Alan Colquitt
Eli Lilly

Amy Grubb
FBI

Kevin Murphy
Colorado State University

Rob Ollander-Krane
Gap, Inc.

Elaine D. Pulakos
CEB

Despite years of research and practice, dissatisfaction with performance appraisal is at an all-time high. Organizations are contemplating changes to their performance management systems, the most controversial of which is whether to eliminate performance ratings. The pros and cons of retaining performance ratings were the subject of a lively, standing-room-only debate at the 2015 Society for Industrial and Organizational Psychology conference in Philadelphia (Adler, 2015). Given the high interest in this topic, this article recaps the points made by the panelists who participated in the debate. The arguments for eliminating ratings include these: (a) the disappointing interventions, (b) the disagreement when multiple raters evaluate the same performance, (c) the failure to develop adequate criteria for evaluating ratings, (d)

Seymour Adler, Aon Hewitt, New York, New York; Michael Campion, Krannert School of Management, Purdue University; Alan Colquitt, Eli Lilly, Indianapolis, Indiana; Amy Grubb, FBI, Washington, DC; Kevin Murphy, College of Natural Sciences, Colorado State University; Rob Ollander-Krane, Gap, Inc., San Franciso, California; Elaine D. Pulakos, CEB, Arlington, Virginia.

This article grew out of a presentation at the 30th Annual Conference of the Society for Industrial and Organizational Psychology in Philadelphia, Pennsylvania (April 23, 2015), titled "Getting Rid of Performance Ratings: Genius or Folly?" (S. Adler, Chair). Authorship order is alphabetical.

Correspondence concerning this article should be addressed to Seymour Adler, Aon Hewitt, 199 Water Street, New York, NY 10038. E-mail: seymour_adler@aon.com

219

the weak relationship between the performance of ratees and the ratings they receive, (e) the conflicting purposes of performance ratings in organizations, (f) the inconsistent effects of performance feedback on subsequent performance, and (g) the weak relationship between performance rating research and practice in organizations. The arguments for retaining ratings include (a) the recognition that changing the rating process is likely to have minimal effect on the performance management process as a whole, (b) performance is always evaluated in some manner, (c) "too hard" is no excuse for industrial–organizational (I-O) psychology, (d) ratings and differentiated evaluations have many merits for improving organizations, (e) artificial tradeoffs are driving organizations to inappropriately abandon ratings, (f) the alternatives to ratings may be worse, and (g) the better questions are these: How could performance ratings be improved, and are we conducting the entire performance management process properly? The article closes with questions organizational members have found useful for driving effective performance management reform.

Background to the Debate (Pulakos)

Despite years of research and practice aimed at improving the performance appraisal and performance management process in organizations, dissatisfaction with the process is at an all-time high. More than 90% of managers, employees, and human resource (HR) heads feel that their performance management processes fail to deliver the results they expected, and many view their current processes as ineffective and/or inaccurate (Corporate Leadership Council, 2012). But even beyond this, study after study has shown that one important component of performance management—that is, the performance review—is dreaded. Formal performance reviews are not only perceived to be of little value, they also can be highly demotivating to even the highest performing employees (Aguinis, Joo, & Gottfredson, 2011; Culbertson, Henning, & Payne, 2013; Rock, 2008). Performance appraisal and performance management are not synonymous, but there is often a strong relation between the two. Although performance management systems often incorporate more frequent and informal feedback, it is common in many organizations to retain and to rely heavily on annual reviews of performance. Pulakos and O'Leary (2011) have argued that the key reason traditional performance management approaches have failed to live up to their promise of enabling performance is that the mechanics of formal systems—how ratings are done, the documentation required, how goals are set—are inconsistent with the goal of providing frequent, credible, and useful feedback about performance. The changing nature of work these days (e.g., semiautonomous teams, remote work, freelancing, temporary work, etc.) also suggests it is time to rethink our fundamental assumptions about approaches to performance management.

One of the main goals of performance appraisal is and always has been simply to achieve high performance by enabling managers to guide

employees to increasingly higher levels of productivity and by motivating employees to do their very best. We have somehow managed to create just the opposite, however: Most employees don't find performance appraisals to be valuable or motivating. Instead, they find performance appraisals and performance management systems frustrating, too bureaucratic, and often not relevant to their jobs. Managers spend a lot of time on formal performance management activities that they likewise believe add little value. The Corporate Leadership Council found that managers and employees spend about 210 and 40 hours, respectively, per year on performance appraisal and performance management activities such as formal goal-setting processes, midyear and year-end reviews, and often extensive rating and calibration processes. For a 10,000-person organization with average salaries, this time, plus the other costs associated with performance management like HR time, training costs, and software, can cost over $35 million dollars annually. This is a very big investment only to disengage employees and potentially undermine performance.

Within the past 2 years, a number of companies have begun contemplating changes to their performance management systems aimed at gaining more value and impact on performance. In fact, we cannot remember a time when so many companies have been uniformly focused on making similar disruptive changes to a major talent system—all at the same time, which is a testament to the ineffectiveness of and dissatisfaction with current performance management processes. However, this has also generated considerable debate about how performance management can and should be changed to gain the most impact on performance. Many organizations have begun changing aspects of their formal performance management systems and introducing training and change management to drive more effective performance management behavior (e.g., real-time feedback, more agile expectations and goals, more collaboration, etc.). Where companies are struggling more is in deciding what aspects of their formal systems to change—with the most controversial and challenging decision surrounding whether to eliminate or retain formal (usually annual) performance ratings. Several high profile companies have opted to eliminate annual performance ratings entirely, such as Eli Lilly, Adobe, and Gap, Inc., but most companies are taking smaller steps, such as introducing more goal-setting and performance check-ins to drive more regular expectation setting and feedback.

The focus on performance ratings, in particular, is noteworthy, especially considering the main point in recent thought pieces on performance management reform (e.g., Pulakos, Mueller-Hanson, Arad, & Moye, 2015; Pulakos & O'Leary, 2011), which is that over 25 years of research have shown that formal system changes have no discernable impact on driving more effective performance management behavior or performance (DeNisi

& Smith, 2014). This may be because formal performance management processes that incorporate or rely on annual performance ratings drive both managers and employees to focus on rating outcomes rather than on how to actually improve and drive effectiveness on the job. Based on the research literature and our collective experience engaging in performance management reform over decades, it is not realistic to think that any intermittent activity-driven evaluation system will necessarily compel more effective performance management behavior or, ultimately, improved job performance. The pros and cons of retaining performance ratings were the subject of a lively, standing-room-only debate at the 2015 Society for Industrial and Organizational Psychology conference in Philadelphia. Given the high interest in this topic, the remainder of this article will recap the points made by the panelists who participated in the debate and advanced a position to keep or to eliminate ratings. The participants in the debate included both academics and practitioners, well-known authors in performance management, highly experienced consultants, and people implementing performance management systems in large organizations in both the private and the public sector. They were allocated to each side of the debate to balance the range of perspectives. They were selected for their knowledge and interest in the topic not because they inherently support one side of the debate or the other. The side for getting rid of performance ratings included Alan Colquitt of Eli Lilly, Kevin Murphy of Colorado State University, and Rob Ollander-Krane of Gap, Inc. The side for retaining performance ratings included Seymour Adler of Aon Hewitt, Michael Campion of Purdue University, and Amy Grubb of the FBI. The debate was moderated by Elaine Pulakos of CEB, who set the background for the debate and provided closing remarks.

Get Rid of Performance Ratings (Colquitt, Murphy, & Ollander-Krane)

There is a long history of research on performance rating and performance appraisal (for reviews, see Bernardin & Beatty, 1984; DeCotiis & Petit, 1978; DeNisi, 2006; DeNisi, Cafferty, & Meglino, 1984; Ilgen & Feldman, 1983; Landy & Farr, 1983; Milkovich & Wigdor, 1991; Murphy & Cleveland, 1991, 1995; Wherry & Bartlett, 1982), and although different reviews highlight different strengths and weakness of the methods that are used in organizations to measure job performance via rating scales or other similar measures, it is fair to say that none of these reviews leads to the conclusion that performance rating is particularly successful either as a tool for accurately measuring employee performance or as a component of a broader program of performance management. Nearly a century of research on performance appraisal (see, for example, Austin & Villanova, 1992) suggests that there is a longstanding history of problems with performance rating and little reason to believe that these problems will be solved in the foreseeable future. The

conclusion that performance rating is not working is not solely an academic one; there is evidence of widespread dissatisfaction with performance rating and related techniques in organizational settings; many large organizations (e.g., Accenture, Deloitte, Microsoft, Gap, Inc., Eli Lilly) have abandoned or substantially curtailed their use of performance appraisal (Culbert & Rout, 2010; Cunningham, 2014).

Most organizations continue to use performance appraisals and ratings, and there is little sign this is changing, at least in the short term (see Lawler, Benson, & McDermott, 2012). Their use affects the distribution of rewards and other outcomes for millions of employees in organizations. The vast majority (89%) of companies link compensation decisions to performance ratings (Mercer, 2013). Performance ratings are the basis for pay-for-performance systems in most organizations; most organizations "pay for performance ratings." It is fair to say that tens if not hundreds of billions of dollars in compensation and rewards are riding on the backs of performance ratings. These ratings can have long and lasting effects on employees' lives and careers in organizations, affecting staffing, promotion, and termination decisions as well as affecting access to other development opportunities. This is serious business, and ratings don't measure up.

The task of accurately evaluating someone's performance is difficult if not impossible. It requires a supervisor to observe the performance of another employee over the course of a year and to collect reports of others' observations of the same employee. The supervisor, usually with little or no formal training, then sifts, sorts, analyzes, weighs, and aggregates this information to make a judgment about the employee's performance. Supervisors must also exclude from consideration other irrelevant information about this individual and any other judgments that may have been made about the individual in the past, and they must suspend any biases or tendencies they possess while making this judgment. Sounds easy enough, right? Judges in professional sports, for whom this is their life's work, have a far simpler task (e.g., count the number of punches each boxer lands), and yet they frequently disagree with each other (see Stallings & Gillmore, 1972). The point of this is what we are asking managers to do is virtually impossible to do well, especially with the frailties of human beings as measurement instruments.

Performance Appraisal: A Failed Experiment

In this section we will argue that it is time to treat performance rating in organizations as a failed experiment. Despite decades of work on methods of improving performance ratings and hundreds of research articles, there is little evidence of meaningful improvement in performance ratings. It is always possible that some future intervention might lead to substantial improvements, but after nearly a century of dashed hopes and unmet

expectations, we believe it is time to call a halt to this sorry business. Scientists rarely like to give up on a problem they have struggled with for so long, but if the ship is sinking, and it is also on fire, it is probably time to head for the lifeboats.

A comprehensive review of all of the shortcomings of performance ratings and of the failed efforts to improve ratings is beyond the scope of this article, but the broad outline of the problems with performance rating and our lack of success in resolving these problems can be summarized under seven headings: (a) the disappointing interventions, (b) the widespread disagreement when multiple raters evaluate the same performance, (c) the failure to develop adequate criteria for evaluating ratings, (d) the weak relationship between the performance of ratees and the ratings they receive, (e) the conflicting purposes of performance rating in organizations, (f) the inconsistent effects of performance feedback on subsequent performance, and (g) the weak relationship between performance rating research and practice in organizations.

Disappointing Interventions

The most common suggestion for improving performance rating in organizations has been to improve rating scales. In particular, substantial efforts have been devoted over several decades to making rating scales clearer and making ratings more reliable and valid by adding valid and detailed behavioral information to rating scales (e.g., behaviorally anchored rating scales, Smith & Kendall, 1963; behavior observation scales, Latham & Wexley, 1977; Murphy, Martin, & Garcia, 1982). The payoff for this effort was judged to be so meager by Landy and Farr (1980), both of whom had themselves published numerous articles on the development of behavior-based rating scales, that they called for a moratorium on further rating scale research. Subsequent reviews (e.g., Murphy & Cleveland, 1995) suggested that behavior-based rating scales might have advantages in terms of their face validity and acceptance by users, but there is little evidence that the substantial body of research on the development and use of these scales made performance ratings in organizations more reliable, valid, or useful.

The second most common intervention has been rater training. This research has followed two main themes. First, numerous studies have relied on the strategy of telling raters about so-called rating errors (e.g., leniency, range restriction, halo) and urging them not to commit these error. Second, numerous studies have developed variations of frame of reference training, designed to ensure that raters adopt a common frame of reference regarding target performance dimensions and performance levels (see, for example, Bernardin & Buckley, 1981; Bernardin & Pierce, 1980; Day & Sulsky, 1995; Pulakos, 1984). On the whole, both types of training "work" in some sense.

If you tell raters not to give ratings that are too high or too low, they are a bit more likely to follow your advice; if you give raters a clearer idea of what the performance dimensions mean and what different performance levels look like, they will show a bit more agreement in their evaluations. However, neither variation on rater training has been successful in markedly improving ratings in organizations (Murphy & Cleveland, 1991, 1995).

Murphy and Cleveland (1995) noted that the interventions that were most commonly used to improve performance ratings were based on the hypothesis that raters lacked the knowledge or the tools to accurately evaluate performance. They suggested that these interventions were probably based on a faulty diagnosis of the shortcomings of performance appraisal in organizations. As we note in a later section, a substantial body of research suggests that there was little real evidence that raters lack the ability to rate accurately and that the likely causes of the failure of performance appraisal are related to raters' motivation and goals rather than to the ability of raters to perform this task.

Disagreement Among Raters

One way to improve ratings might be to collect evaluations of performance from multiple raters rather than relying on the judgment of any single rater. Unfortunately, there is consistent evidence that raters do not agree in their evaluations of ratees. Disagreement is probably more substantial when raters differ in terms of their relationship with the persons to be rated (e.g., supervisors vs. peers), but even when raters are at the same level in an organization, they often do not agree in their evaluations (Facteau & Craig, 2001; Harris & Schaubroeck, 1988; Heneman, 1974; Murphy, Cleveland, & Mohler, 2001; Viswesvaran, Ones, & Schmidt, 1996). There have been intense disagreements about precisely how disagreements among raters should be interpreted and about the more general implications of these disagreements for the psychometric quality of ratings (e.g., Murphy & DeShon, 2000; Ones, Viswesvaran, & Schmidt, 2008), but there is little controversy about the fact that raters do not show the level of agreement one might expect from, for example, two different forms of the same paper-and-pencil test. Viswesvaran et al. (1996) suggest that performance ratings obtained from two separate sets of raters in organizations are likely to show correlations in the .50s, hardly the level one would expect if ratings were in fact good measures of the performance of ratees.

In theory, it is possible to obtain ratings from multiple raters and pool them to eliminate some types of interrater disagreement. However, in practice this is often difficult. Rating systems that obtain information from multiple raters (e.g., 360 degree feedback programs) often sample raters from different levels of the organization or raters from outside of the organization

(e.g., customers), and there are good reasons to believe that differences in roles and perspectives will introduce systematic disagreements among raters that might not be readily resolved by simply pooling ratings (Murphy, Cleveland, & Mohler, 2001).

Weak Criteria for Evaluating Ratings

Suppose N raters each evaluate the performance of k subordinates. How do you know whether the ratings they provide do a good job measuring the performance of the people being rated? If multiple raters evaluate the same ratees, interrater agreement measures provide one criterion, and as noted above, the data regarding interrater agreement are far from encouraging. The more common approach has been to evaluate performance ratings in terms of the presence or absence of one or more so-called "rater errors."

Seventy-five years ago, Bingham (1939) identified what would come to be called "halo error" in ratings. Other rater errors, including leniency or severity and range restriction, were soon identified. There is a substantial research literature that analyzes rater error measures, particularly measures of halo (Balzer & Sulsky, 1992; Cooper, 1981a, 1981b; Murphy, 1982; Murphy & Anhalt, 1992; Murphy & Balzer, 1989; Murphy, Jako, & Anhalt, 1993; Nathan & Tippins, 1989; Pulakos, Schmitt, & Ostroff, 1986). On the whole, rater error measures are deeply flawed criteria. Different measures of halo, leniency, and the like do not necessarily agree. More important, all of these measures are based on completely arbitrary assumptions about how performance is distributed. Suppose, for example, that I give ratings of 4 to 5 while you give ratings of 2 or 3. This could mean that I am lenient, that you are severe, or that the people who work for me are performing better than the people who work for you are, and in principle there might be no way to determine which explanation is correct. Suppose that I rate my subordinates on four performance dimensions (e.g., planning, written communication, attention to detail, and receptiveness to customer feedback) and that the average intercorrelation among the ratings I give is .50. Is this too high, or is it too low? How do you know how high these correlations *should* be? Maybe I am committing halo error, or maybe the people who are good at planning really are also good at oral communication. Outside of artificial laboratory environments, there is usually no a priori way of determining whether the correlations among ratings is too high, too low, or just right.

Perhaps the most important critique of halo error measures as criteria is the paucity of evidence that ratings that do not exhibit what we judge to be halo, leniency, or the like are better measures of job performance than ratings that are suffused with "rater errors." Earlier, we noted that one popular form or training instructed raters not to make halo errors, not to be lenient, and so forth. It is not clear that removing halo, leniency, and the like makes

performance ratings better measures of performance, and it is even plausible that efforts to suppress these so-called errors make performance ratings worse (Murphy, 1982; Murphy, Jako, & Anhalt, 1993).

Borman (1977) argued that under some circumstances, it might be possible to assess the accuracy of performance ratings. The prototypical rating accuracy study would collect ratings under standardized conditions in which all raters viewed the same performance (e.g., using videotaped vignettes) and in which the pooled judgments of multiple experts could be used as a criterion for evaluating accuracy. There is a vibrant literature dealing with the measurement and interpretation of rating accuracy (e.g., Murphy & Balzer, 1989; Sulsky & Balzer, 1988), and in general, this literature leads to three conclusions: (a) Rater error measures are virtually useless as indices of rating accuracy; (b) there are multiple ways of defining accuracy in rating, and the different accuracy indices are often essentially unrelated to one another; and (c) outside of tightly controlled experimental settings, it is prohibitively difficult to measure accuracy directly in the field.

In sum, there are two classes of measures we can use in field settings to evaluate ratings: rater agreement measures and rater error measures. Rater agreement measures tell a sorry story, but it is one whose full implications are not fully understood (Murphy & DeShon, 2000; Viswesvaran et al., 1996), whereas rater error measures have no clear value as criteria.

Contextual Effects on Ratings

Although the most widely cited conclusion of Landy and Farr's (1980) review of performance appraisal research was that rating scales probably do not have a large effect on the quality of rating data, perhaps a more important conclusion was that there are a number of variables other than job performance that have a clear influence on performance ratings (see also Murphy, 2008; Scullen, Mount, & Judge, 2003). For example, a number of authors (e.g., Grey & Kipnis, 1976; Murphy & Cleveland, 1991, 1995) have suggested that contextual factors, ranging from broad societal factors (e.g., the state of the economy and the labor market) to the organizational context in which ratings are collected (e.g., the climate and culture of the organization), are likely to affect performance ratings. Other authors (e.g., Aguinis & Pierce, 2008; Ferris, Munyon, Basik, & Buckley, 2008; Judge & Ferris, 1993) have noted the importance of the social context of rating such as social power, influence, leadership, trust, social exchange, group dynamics, negotiation, communication, and other issues.

One of the nonperformance determinants of performance ratings that has been most extensively studied can best be described as the political use of rating (Cleveland & Murphy, 1992; Longenecker, Sims, & Gioia, 1987; Murphy & Cleveland, 1995; Tziner & Murphy, 1999; Tziner, Prince, & Murphy,

1997). Research on the political aspects of performance appraisal suggests that raters pursue a variety of goals when completing performance appraisals and that these goals substantially influence the ratings they give (Murphy, Cleveland, Skattebo, & Kinney, 2004). Murphy and Cleveland (1995) suggest that these goals include (a) task performance goals—using performance ratings to influence the subsequent performance of ratees; (b) interpersonal goals—using appraisal to maintain or improve interpersonal relations between the supervisor and the subordinates; (c) strategic goals—using appraisal to increase the supervisor's and/or the workgroup's standing in the organization; and (d) internalized goals—goals that are the results of raters' beliefs about how he or she should evaluate performance (e.g., some raters want to convey the image of being a hard case, with high performance standards, like the professor who announces at the beginning of a class that he or she rarely gives "A" grades). On the whole, the goals raters pursue tend to push those raters in the direction of the "Lake Woebegone Effect" (i.e., giving high ratings to just about all employees). High ratings will maximize rewards (thereby potentially maximizing motivation), will preserve good relationships between supervisors and subordinates, and will make the supervisor look good (a supervisor whose subordinates are performing poorly will seem ineffective). In most organizations, there are few concrete rewards for giving accurate ratings and few sanctions for giving inflated ratings.

In our experience working with performance appraisal in real-world settings, a number of reasons are often cited by supervisors or managers for systematically manipulating the ratings they give. These are summarized in Table 1.

Conflicting Purposes

Ratings are used for many purposes in organizations (e.g., to determine promotions and salary increases, to evaluate developmental needs, to serve as criteria for validating tests and assessments, to provide documentation for legal purposes), and these purposes often come into conflict (Cleveland, Murphy, & Williams, 1989; Murphy & Cleveland, 1995). Fifty years ago, Meyer, Kay, and French (1965) noted the incompatibility of using performance appraisal simultaneously to make decisions about rewards and to provide useful feedback and suggested that the different uses of performance appraisal should be separated by creating different evaluative mechanisms for rewards versus feedback. Murphy and Cleveland (1995) suggested that uses of performance appraisal to highlight differences between people (e.g., salary, promotion, validation) are fundamentally at odds with uses of appraisal to highlight differences within persons (e.g., identifying developmental strengths and weaknesses).

Table 1. Commonly Cited Reasons To Manipulate Performance Ratings

Overrate
- Get more money for employees.
- Encourage employees.
- Avoid painful outcomes for him/herself by not rating someone low (e.g., the work required by a disciplinary process).
- Avoid hurting employees now or in the future by giving them low ratings.
- Avoid confrontation with an employee by giving him or her a rating lower than he or she expect.
- Get rid of a bad employee so he or she is marketable elsewhere within the company.
- Make themselves look good. Low ratings might imply the supervisor is not doing a good job.
- Get employees access to opportunities that are dependent on higher performance ratings (e.g., development opportunities).

Underrate
- Make an example out of an employee. A supervisor may rate employees lower than they deserve to punish them or teach them a lesson.
- Get rid of an employee. Supervisors may rate someone lower than he or she deserves simply to create the necessary documentation and evidence to fire someone they don't like.
- Send a message. A supervisor may rate an employee lower simply to send the message that the employee should leave.

Cleveland et al.'s (1989) survey suggested that most organizations used performance appraisals for multiple conflicting purposes and that this attempt to satisfy incompatible goals with performance appraisal was one source for the continuing evidence of dissatisfaction with performance appraisal in organizations. Some goals might be compatible. For example, use of performance appraisal for salary administration is not necessarily incompatible with using performance appraisal as a tool for validating selection instruments (both uses focus on differences between ratees). However, even when there is no underlying conflict among uses, the particular dynamics of one use of appraisals (e.g., for determining raises) might limit the value of the appraisal process for other uses (e.g., the range restriction that is likely when most ratees receive high ratings might limit the value of appraisals as criteria for validating selection tests).

Feedback Is Not Accepted or Acted On

One of the core assumptions of virtually all performance management programs is that if people receive feedback about their performance, they will be both motivated and empowered to improve (assuming that appropriate organizational supports and incentives are in place to support this improvement). This assumption does not hold up well, at least in the context of

performance appraisal. There is evidence that employees dislike giving or receiving performance feedback (Cleveland, Murphy, & Lim, 2007). The performance feedback they receive is often inconsistent and unreliable (Murphy et al., 2001). Even if feedback is reasonably accurate, the effects of feedback are inconsistent, sometimes improving subsequent performance, sometimes making things worse, and sometimes having little discernible effect (Kluger & DeNisi, 1996).

Accurate feedback requires accurate evaluations of performance, and there is little good reason to believe that performance appraisals in organizational settings provide this sort of accuracy. However, even if performance feedback were accurate, there is little reason to believe that it would be consistently accepted and acted on. It is well known that peoples' self-ratings of performance are consistently higher than ratings from supervisors or subordinates (Harris & Schaubroeck, 1988; Heneman, 1974).

Differences in self-ratings versus others' ratings are not a shortcoming of performance appraisal per se but rather a reflection of broadly relevant processes in the way we understand our own behavior versus the behavior of others. For example, there is a fundamental attribution error (Ross, 1977) that makes us likely to attribute our own successes to internal factors (e.g., skill, effort) and our own failures to external ones (e.g., luck, lack of opportunity) but to make the opposite attributions when others succeed or fail. In other words, we are likely to take credit when we succeed and to avoid blame when we fail but to see others' success and failure in starkly different terms, making it likely that we will view our own performance favorably even when we have in fact performed poorly. As a result, feedback that is unbiased and accurate will sometimes seem harsh and unfair.

Performance appraisal and performance management systems are often built on assumptions that may not be warranted. The first assumption is that employees want to be rated. Managers assume employees want to "know where they stand," and they assume this means employees want some form of rating or relative comparison. This makes sense. People compare themselves with others. This is the basis for social comparison, equity, and gaming theory (see Adams, 1965; Bryson & Read, 2009; Festinger, 1954). However, this doesn't mean that people want to be compared with others by third parties. It is also true that people generally think they are above average (Meyer, 1980), so most feel they will benefit from these ratings and comparison. Although employees may say they want to be rated, their actions suggest otherwise. Barankay (2011) showed that a majority (75%) of people say they want to know how they are rated (or ranked in this case), but when given the opportunity to choose between a job where their performance was rated against others and one where it was not, most choose the work without ratings.

The second assumption is that performance ratings motivate employees—those getting poor ratings are motivated to improve, and those getting top ratings are motivated to keep them. This assumption comes from labor economics and tournament theory (see Connelly, Tihanyi, Crook, & Gangloff, 2014). Research suggests this assumption does not always hold up. Although some research finds ratings improve motivation and performance (see Blanes i Vidal & Nossol, 2009), other research suggests they can hurt both motivation and performance, especially in collaborative work environments and when ratings are tied to differential rewards (see, e.g., Casas-Arce & Martinez-Jerez, 2009). On balance, ratings seem to hurt more than they help.

Performance feedback represents the ultimate lose–lose scenario. It is extremely difficult to do well, and if it was done well, the recipients would be likely to dismiss their feedback as inaccurate and unfair. It is no wonder that supervisors and subordinates alike approach performance appraisal with a mix of skepticism and unease.

Weak Research–Practice Relationships

Decades of research and hundreds of articles, chapters, and books dealing with performance appraisal seem to have little impact on the way performance appraisal is done or the way information from a performance appraisal is used in organizations (Banks & Murphy, 1985; Ilgen, Barnes-Farrell, & McKellin, 1993). Banks and Murphy (1985) suggest that this is due in part to a lack of fit between the questions that are of greatest interest to researchers and the problems practitioners need to solve. For example, a significant portion of the performance appraisal studies published in the 1980s and 1990s were concerned with cognitive processes in appraisal (e.g., attention allocation, recognition, and recall of performance-related information; DeNisi, 2006). For psychologists, these are very important and interesting questions, but they are not questions that are necessarily relevant to the practical application of performance appraisal in organizations. It is worth noting that the lack of a solid research–practice relationship is not solely the result of academic disengagement from the realities of the workplace. Murphy and Cleveland (1995) presented a list of practical applications and research-based suggestions for improving performance appraisal, and there is little evidence that these were ever taken up and applied by practitioners.

One potential explanation for the failure of performance appraisal research to address the problems of performance appraisal is that these problems may be more intractable than we want to admit. Several researchers (e.g., Banks & Murphy, 1985; Murphy & Cleveland, 1995) have suggested that the shortcomings of performance appraisal are not an indication that raters *cannot* effectively evaluate job performance but rather an indication

that raters and more generally organizations do not *want* to solve the problem. For example, organizations do not appear to do anything to reward supervisors who do a good job with performance appraisal or to sanction those who do a bad job. Accuracy in performance appraisal is likely to make subordinates unhappy and demotivated, and it has the potential to make subordinates look bad. Indeed, if a manager's job is to help get the most out of his or her subordinates, a case can be made that performance appraisal (even at its best) interferes with that job and that a smart manager will do all he or she can to subvert and ignore the appraisal system (Murphy & Cleveland, 1995).

Institutional theory provides some insight into the difficulty of changing performance appraisal and performance management practices in organizations by means of academic research. Performance management and rating practices get embedded and become hard to change. These practices get institutionalized in organizations, and companies copy and adopt them because they are "the standard." Companies also imitate and copy them in order to achieve legitimacy (Pfeffer, 2007; Scott, 1995). Social relationships created among organizational members (especially senior leaders who are members of each other's boards) also contribute to this conformity and imitation behavior. General Electric's (and Jack Welch's) adoption of forced ranking systems in the 1990s is a good example.

Additional Negative Consequences of Ratings

Finally, there are several additional potential drawbacks of ratings that were not mentioned above. First, ratings sometimes bring us unpleasant news about diversity differences (see meta-analysis of race differences by McKay & McDaniel, 2006). Second, they are the focus of many lawsuits on topics such as promotion, compensation, and terminations. Third, they are often not helpful when you really need them, such as in terminations and downsizing, when you find they have not sufficiently documented poor performance or have too much skew to be useful for decision making.

Get Rid of Ratings but Tread Carefully

Most companies stay with ratings because they don't know what else to do and because ratings are inputs to so many other processes, especially rewards. How will organizations distribute rewards and make promotion and termination decisions without performance ratings? Rest assured these are all solvable problems. These decisions can all be made without performance ratings and will probably be better without them. However, it is also important to realize there are also some significant philosophical questions wrapped up in a decision to abandon performance ratings that are beyond the scope of this article. Most organizations

have pay-for-performance philosophies and see themselves as meritocracies. In most companies, performance ratings are the primary way of determining merit. Well over 90% of companies tie pay to performance to at least some extent, and 82% link individual performance to compensation (Mercer, 2013). How will organizations determine merit without ratings? How do organizations design performance-based rewards systems without individual performance ratings? The answer is they don't—at least not individually based reward systems. Companies should not abandon performance ratings without wrestling with these issues. There are effective solutions to these problems as well, but we will save these arguments for another day (or article).

Why Getting Rid of Performance Ratings Is a Bad Idea (Adler, Campion, & Grubb)

Performance evaluation is a complex, difficult, and often unpleasant part of management but is essential in some form if an organization wants to be a meritocracy. In this section, we argue that eliminating performance ratings is probably not the solution for meaningfully addressing the widespread dissatisfaction with performance management. We start by clarifying the point of the debate. Then we argue that performance is always rated in some manner, and being difficult to do is not an excuse. We then summarize the merits of ratings (no pun intended) and the value of differentiated evaluations. We suggest that artificial tradeoffs are driving organizations to abandon ratings. We then describe the alternatives to performance ratings, most of which are even less pleasant. Finally, we provide some suggestions for the improvement rather than the abandonment of performance ratings.

Performance Ratings ≠ Performance Management

Let's start with a clarifying observation on the terms of this debate because some of the issues described above have ranged outside the scope. When addressing the question of using or discarding performance ratings within the larger context of performance management and all that ails it, it may stand out as odd (and possibly unproductive; see Landy & Farr, 1980) to have the conversation focused on just a single element of performance management: the use of performance ratings. There are a myriad of complex and interwoven issues that need untangling and restringing when it comes to the broader and more significant task of effectively managing an employee's performance (e.g., Banks & Murphy, 1985; Curtis, Harvey, & Ravden, 2005). But instead, our field seems to have focused on perhaps the least meaningful element of performance management but one that is most comfortable for I-O psychologists—measurement, also known as the ratings. A prime illustration of this disproportionate focus is the extensive summary of the 75-year

234 SEYMOUR ADLER ET AL.

history of empirical literature on ratings presented above by the "antiratings" side.

To be sure, HR professionals, leaders, and I-O psychologists all widely agree that how performance management—including the assignment of performance ratings—is practiced within the business environment is badly in need of improvement. However, we also need to recognize that, fundamentally, the goal of an organization is to perform well and that this goal necessarily requires the individuals within the organization to perform well. The evaluation of individual performance then becomes an indispensable component of measuring and achieving organizational success. To us, this logic, however painful, is inescapable.

Let's be clear, then, about the center of the debate. This is not about whether there is value in core aspects of performance management (e.g., goals, regular constructive feedback, development focus, etc.). We all agree with that. The issue is whether we should continue to attempt to collect a psychometrically meaningful quantitative performance index of some sort.

Performance Is Always Evaluated

Whether called ratings, standards, or judgments, leaders and employees make decisions and evaluations about the performance, skills, attributes, and assets of those around them all the time. Judgments that employees "don't have it," "have it," or "knock it out of the park" are made every day in organizations. Dozens of conceptualizations of what can and should be measured within an individual performance system have been put forth in the literature and in practice (five-, four-, three-, two-point rating scales; competency based, production based, objective based, task based, behavior based, etc.; e.g., Benson, Buckley, & Hall, 1988; Bommer, Johnson, Rich, Podsakoff, & MacKenzie, 1995; Borman, 1979; Catano, Darr, & Campbell, 2007; Fay & Latham, 1982; Goffin, Jelley, Powell, & Johnston, 2009; Jelley & Goffin, 2001; Johnson, 2001; Latham, Fay, & Saari, 1979; Thompson & Thompson, 1985; Tziner & Kopelman, 2002). Irrespective of the manner in which evaluations happen, judgments are made and decisions about rewards and roles in the future are based on those judgments. For corporate success, these judgments should be grounded in something tangible and documented and, ideally, something that furthers the performance of the organization as a whole. The organization needs to have a common language, common perceptions of what is important, and a starting point of where the population is in different skills, achievements, and attributes in order to manage individual and collective performance forward. If there is no starting point metric, how can an organization gauge its progress? These judgments, standards, and evaluations are all essentially what we mean by the term "ratings," whether or not they are labeled as such.

For those organizations that promote a "development culture" and feel they need to move away from using ratings to do so, we remind them that judgments of where people are in their professional growth are still being made. Development conversations start with a picture of where employees are now, where they are going, and the actions needed to close that delta. The "where they are now" is a judgment on some evaluation scale. Although assigning or being assigned a "rating" may be uncomfortable for some (Rock, Davis, & Jones, 2014), the science of behavioral change based on control theory also tells us that some discomfort, some feedback on the gap between the "now" state and the desired future state is required to stimulate behavior change and development (Carver, Sutton, & Scheier, 2000). Indeed, in organizations with a genuine development culture, detailed evaluative ratings feedback is not the threatening beast as characterized in recent popular press. In organizations with a genuine developmental culture, employees—and especially high-potential employees and particularly Millennials—manage their own development and proactively seek out feedback on areas of growth and improvement. Those areas can only be identified through the measurement of performance evaluations against established standards, a process otherwise known as rating.

Reflecting the ubiquity and inescapability of evaluative judgments of performance, we should recognize that many—perhaps most—organizations that have trumpeted their "daring" elimination of ratings, have actually done no such thing. Many continue to have managers enter a number into the performance management system to represent a recommended percentage increase in salary or a recommended percent-of-target-bonus award. These entries are intended to quantitatively reflect prior year performance—we would call that activity "rating." Other organizations claiming to have eliminated ratings continue to use some sort of traditional rating scale but have only eliminated imposing a forced distribution due to employee reactions (Blume, Baldwin, & Rubin, 2009; Chattopadhayay & Ghosh, 2012; Schleicher, Bull, & Green, 2008). Others have simplified their rating process, typically reducing the number of dimensions rated (e.g., fewer individual goals rated, a one-item global rating of the "how" of performance in lieu of rating a long list of individual competencies). Still others have simply changed the wording of the dimensions rated. For example, some organizations have substituted the term "overall impact" for "overall performance." One more creative example is Deloitte's recent adoption of a three-item scale (a fourth item measures promotion potential) instead of more traditional performance rating formats (Buckingham & Goodall, 2015). A sample item is "Given what I know of this person's performance, I would always want him or her on my team." Many of these organizations have made laudable improvements in

aspects of performance management—but they surely did not "get rid of ratings."

"Too Hard" Is No Excuse for I-O Psychology

Many in the field seem to have turned the conversation about performance management into a lament that "performance is too *hard* to measure," which therefore leads to the suggestion that we abandon the science and practice of defining and measuring what "performing well" means. To this we say, differential psychology is, and always has been, a core—arguably, *the* core— discipline of our field (Chamorro-Premuzic, 2011). Many (most?) I-O psychologists spend their careers attempting to quantify how people differ in attributes as varied as knowledge, skill, engagement, safety orientation, narcissism, and—yes—job performance. We have a well-developed set of psychometric principles and tools that allow us to evaluate and improve the reliability and validity of our measures.

Admittedly, job performance is a complex construct (Campbell, McCloy, Oppler, & Sager, 1993). Its definition varies across organizational contexts and roles. There are almost as many ways that job performance has been measured as there are jobs (Aguinis, 2013). When assessed, the psychometric characteristics of measures of job performance range widely, with most measures showing weak-to-moderate reliability and construct-oriented validity; this is well documented in the prior section of this article. But a history of measurement challenges is not an excuse for abandoning the differential psychology of performance for the "armchair" assignment of idiosyncratic, vague qualitative descriptors of performance. Giving up on the effort to improve the differential measurement of performance should simply not be an option for either I-O scientists or practitioners. Note also that for many roles across a large segment of the workforce, including for those in government and other organizations, there is actually no real option of abandoning ratings, whatever their flaws; ratings are mandated by law or regulation (e.g., civil service merit-principle requirements) or traditional practice (e.g., think teacher evaluations). It is therefore critical that our field continues to explore supportable ways to rate individual performance.

What Are the Merits of Ratings?

Are there *any* merits to having a psychometric index of job performance? Although not perfect, there is evidence supporting the reliability and validity of performance ratings. For example, meta-analytic research shows at least modest reliability for performance ratings (Rothstein, 1990; Viswesvaran et al., 1996); if ratings can be pooled across many similarly situated raters, it should be possible to obtain quite reliable assessments. Despite the current debate, there is probably general agreement among I-O psychologists that

performance ratings, if done well, share at least some meaningful true score variance with actual job performance. If collected for research purposes, performance ratings have been shown to be a useful criterion in thousands of validation studies of selection procedures in the history of our profession (Schmidt & Hunter, 1998) and are recommended in our testing principles (Society for Industrial and Organizational Psychology, 2003).

Applied psychometrics is what I-O psychologists are good at, and it brings great value in other domains. As examples, consider how well it helps us measure the knowledge and skills of candidates in personnel selection or measure the attitudes and beliefs of employees in engagement surveys. Would we really be better off not to at least attempt to generate a psychometric index to add to the performance management system? Would we really be better off with just unreliable qualitative data?

Performance rating that depends on the subjective judgments of one or more raters is not necessarily the *preferred* method of evaluating job performance; it is the *only feasible* method for attacking this problem. Objective measures (e.g., absenteeism, production counts) are almost always deficient criteria (Landy & Farr, 1980). Frequent informal feedback, something that is often suggested by proponents of performance management (e.g., Aguinis, 2013), would be difficult to validate in any systematic way. Establishing the job relatedness of informal feedback would be difficult, and establishing the fairness and freedom from bias of an informal evaluation system would be even more daunting. Like it or not, performance rating by supervisors or by some other source that is knowledgeable about the job (e.g., peers) remains our best bet for obtaining useful measures of performance.

The Value of Differentiated Evaluations

Research has consistently shown that differentiating rewards and recognition based on quantified performance differences actually contributes to enhanced individual and organizational performance. Four closely related lines of research have consistently shown the following:

- Strong performers are differentially attracted to organizations that recognize individual contributions (e.g., Menefee & Murphy, 2004). For instance, Trank, Rynes, and Bretz (2002) demonstrated that high achievers have a stronger preference for organizations that apply individual rather than group pay systems, emphasize praise and recognition for individual accomplishment, and employ fast-track promotion practices for top performers. Not surprisingly, strong performers consistently choose organizations that have a reputation for pay-for-performance systems over fixed salary compensation systems (e.g., Cadsby, Song, & Tapon, 2007).

- Strong performers will leave organizations that do not differentially reward strong performance (e.g., Allen & Griffeth, 2001). As an example, Trevor, Gerhart, and Boudreau (1997) showed that low salary growth is related to high turnover among top performers, relative to low performers.
- Weak performers will leave organizations that strongly reward high performance (e.g., Lazear, 1986). Corresponding to the finding regarding low salary growth, Trevor, Gerhart, and Boudreau (1997) showed that high salary growth is related to higher turnover among low performers but to low turnover among top performers.
- Weak performers are more likely to stay with an employer when pay-for-performance differentiation is weak (e.g., Harrison, Virick, & William, 1996). For example, Williams and Livingstone (1994) showed that the relationship between performance and turnover is significantly weaker when rewards are not linked to performance than when they are linked to performance.

Recent CEB Corporate Leadership Council (2014) research, based on a sample of over 10,000, showed that Millennials are particular motivated by relative, rather than absolute, performance feedback, compared with the broader employee population. This research shows that, in contrast to the myth that characterizes Millennials as more collaborative, they are actually more competitive and have a strong desire for performance evaluation systems that provide clear differentiation from peers. One conclusion, then, is that the importance of performance differentiation may indeed grow as Millennials become the core segment of the working population.

Artificial Tradeoffs Are Driving Organizations To Abandon Ratings

If the benefits of measuring differences in performance and rewarding differentially based on those performance differences are so clear, why have some small number of organizations abandoned the attempt to measure individual differences in performance? We believe this recent trend (if there is one) stems at least in part from a set of falsely dichotomized choices:

Quantification versus humanization. There is a belief that attaching a rating to the quality of an individual's job performance somehow dehumanizes that individual, reducing him or her to a number. That indeed is a risk. How often have we heard employees referred to as "*he's* a 2" or "*she's* a 4" instead of the accurate and appropriate "her *performance* last year was a 4; it exceeded the established goals and expectations." Quantification can be dehumanizing but does not have to be. Managers need to learn and apply the language of effective performance management; indeed, the impact of feedback on performance improvement is strongest when that feedback is focused on performance and on partnering with the employee on the means

of enhancing performance rather than on labeling performance as an inherent quality of the performer. The fact that rating labels are misused is not a reason to avoid ratings. There is a large research literature in our field on how to improve reactions to performance evaluation (e.g., Bobko & Colella, 1994; Cawley, Keeping, & Levy, 1998; Dipboye & Pontbriand, 1981; Dorfman, Stephan, & Loveland, 1986; Giles & Mossholder, 1990; Ilgen, Peterson, Martin, & Boeschen, 1981; Pichler, 2012; Silverman & Wexley, 1984).

Bad performance management versus no performance ratings. In many—perhaps even most—organizations, performance management practices are not well-implemented, as both sides of this debate concede. As argued by the antiratings side above, appraisal forms are often overengineered in an attempt to satisfy multiple organizational objectives (e.g., rewards, developmental planning, succession, legal and regulatory compliance, etc.). Managers are neither selected nor adequately trained to manage performance in many instances (Murphy, 2008). Rating guidelines are often confusing, poorly understood, and inconsistently applied across raters and, within raters, across ratees. Organizations at times require ratings to fit distribution curves, which may help reduce leniency (Scullen, Bergey, & Aiman-Smith, 2005) but may not represent the true distribution of performance (Aguinis & O'Boyle, 2014, but cf. Beck, Beatty, & Sackett, 2014). All these flaws can be and in many organizations are being addressed (see the example of Cargill in Pulakos et al., 2015). There is no reason to add to bad performance management practice by eliminating a core area of managerial accountability.

Ratings versus conversations. The value of active, ongoing, constructive feedback, performance strategy, and coaching conversations between managers and employees has been long recognized (Meyer et al., 1965). No one disputes that. However, there is no reason to believe that eliminating systematic, contextually bounded, and documented performance evaluations will result in an increase in the frequency or quality of those conversations. Where is the evidence that the only reason managers do not conduct more frequent, more engaging, and more impactful conversations is that they are too busy filling out appraisal forms? Opponents of ratings cite a recent Korn Ferry study that indicated, across companies, the competency consistently rated lowest (67 of 67 on the Lominger list!) is growing talent (Rock et al., 2014). It would seem to be naïve to think that, relieved of the burden of ratings and without the "crutch" of a structured feedback tool, managers will somehow overcome this weakness and consistently engage in positive and impactful conversations. Indeed, one important mechanism for assuring that quality ongoing performance conversations occur—and acknowledging Murphy's (2008) argument that it is more a matter of manager motivation than skill—is by setting goals as well as evaluating and rewarding managers for how effectively they manage the performance of their own subordinates.

Feedback versus motivation. Feedback—if provided destructively—can indeed be a demotivator (Rock, 2008). Poorly communicated feedback is an aspect of poor leader performance. Effective feedback is motivating. Research has shown that feedback can stimulate the development of more effective performance strategies. Brockner (1988) has shown that people differ in behavioral plasticity and that, although some do have difficulty internalizing feedback, others demonstrate meaningful behavioral change when provided with feedback. Further, opponents of ratings are fond of citing Dweck's (2006) mindset work in support of eliminating ratings. However, that research actually shows that it is not performance feedback per se that is the issue; it is the interpretation of that feedback that is key. Those with a learning mindset benefit from performance feedback; it is just that they use that feedback constructively to improve their skills and ultimately their performance. Effective performance management requires the constructive communication of specific, actionable feedback, not the avoidance of feedback.

In other words, we are arguing that organizations can have it all—*if* they apply our science and effective performance management practice. Managers can quantify individual differences in performance and treat their employees as partners in performance management. They can design administratively simple but impactful performance management practices and train managers to apply those practices to drive better individual and collective performance. They can learn to validly assess performance and, based on those assessments, engage employees in frequent, constructive dialogues in ways that are empowering, motivating, and inspiring to improve performance and employee engagement.

What Are the Alternatives to Performance Ratings?

Let us consider the consequences of living without ratings. Consider the potential answers to questions such as the following:

- Will performance measurement be more precise without a rating? This seems unlikely because the only measurement will be narrative data, which are inherently less reliable.
- Will merit-based management be improved? With no measure of merit other than narrative comments, it is hard to imagine how merit-based management, like pay increases and promotions, would be improved.
- Will employees be more motivated to work hard? Antagonists of ratings argue that eliminating ratings will indeed be more motivating, but that is yet an untested hypothesis. It is certainly true that most employees prefer not to be rated (Cleveland et al., 2007), but achieving higher ratings (and avoiding low ones) is very motivational.

- Will lawsuits be easier to defend? Without any measure of performance, it would likely be more difficult to defend HR decisions based on performance. However, the existence of diversity differences in performance ratings is sometimes used in an attempt to establish a claim of discrimination. In addition, if the organization does not use the performance ratings consistently and fairly to determine HR decisions, this might create more of a liability than a defense. So the answer to this question may be uncertain.

- How would we make compensation decisions? Across-the-board increases serve little value in motivating higher performance. Step increases based on time in grade are similarly limited. Standard of living increases are the least desirable way to allocate pay increases because they imply that increases are somehow required of the organization to keep pace with the cost of living and not related to individual performance. Organizations can only give increases when there is a change of job or responsibilities. This is not uncommon, especially in small organizations, but it not as useful for improving performance because of the limited availability of higher level jobs. Finally, but by no means the last possibility, management could use subjective judgment based on the narrative-only performance appraisal. This would maintain the appearance of a link to performance but would likely be difficult to justify to employees getting smaller pay increases.

- How would we make promotion decisions? We could rely on selection tools and not past performance. However, past performance in the same company in similar jobs is often considered to be an excellent predictor of performance after promotion (Beehr, Taber, & Walsh, 1980; DeVaro, 2006), and employees have a strong expectation that past performance should be a factor considered in promotion (De Pater, Van Vianen, Bechtoldt, & Klehe, 2009). We could just use seniority, like most union contracts require. This may be viewed as fair by some employees, especially in environments with low trust in management, but it does not ensure that the highest performers or the most skilled are promoted. We could rely on succession management systems, which refer to management planning discussions of promotion potential, but the results of these are usually not shared with employees and thus do not serve the purposes of motivating performance or ensuring fairness perceptions. We could use promotion panels to read the narrative appraisals and recommend promotions. This is common in some settings, especially in the public sector, and has many advantages, such as considering a wide range of inputs, multiple decision makers, and a consistent process, but this is very costly in terms of management time. Finally, to complete this nonexhaustive list, we could qualitatively consider past performance based on the narrative-only appraisals. Again, this would maintain the appearance of a link to performance but would likely be

difficult to justify to the employees who are not promoted without any metric of performance.

The Path Forward

Perhaps the better question than whether or not to get rid of ratings is this: How could performance ratings be improved? Many choices of rating scales and expected distributions of ratings might help, depending on the context and culture. A recent trend with considerable merit is the formal use of "calibration," which refers to meetings among managers for the purpose of justifying and comparing the proposed performance ratings of employees to ensure their accuracy and comparability across business units (Catano et al., 2007; Church, 1995; DeNisi & Pritchard, 2006). Another increasing trend is competency modeling, which involves (in part) the explicit establishment of behavioral descriptions of performance expectations, at each rating level, for important job competencies (M. A. Campion et al., 2011; Catano et al., 2007; Halim, 2011). Yet another trend worth considering is the use of 360 feedback as an input to performance management (Bracken & Rose, 2011; M. C. Campion, Campion, & Campion, 2015; Gilliland & Langdon, 1998; Latham, Almost, Mann, & Moore, 2005; Reilly & McGourty, 1998). Multirater 360 feedback in the form of either formal ratings or informal input may be especially relevant as work becomes more team based, less hierarchical, and more customer focused, as other changes result in additional parties having performance information on an employee, and as work arrangements change (e.g., remote work) such that the manager has less of an opportunity to observe the work of the employee. It has many potential advantages, such as increased reliability, reduced bias, and reduced leniency (e.g., Aguinis, Gottfredson, & Joo, 2013; Antonioni, 1994; Flint, 1999; London & Wohlers, 1991; Smither, London, & Reilly, 2005). Simplifying the cognitive demands and reducing, through simplification of the process and forms, the motivational barriers for managers is also likely to be helpful (Efron & Ort, 2010). Finally, performance review panels might be used, which would have similar advantages and disadvantages as promotion panels (Catano et al., 2007; Church, 1995; Gilliland & Langdon, 1998; Kozlowski, Chao, & Morrison, 1998; Vance, Winne, & Wright, 1983; Werner & Bolino, 1997).

In our view, it is the *consequences* of leaders assigning and communicating evaluative judgments expressed as ratings that matter the most, not the ratings themselves. Leaders are the front line of performance management. The stronger the promotion process to select leaders who value development, and the stronger the formal and informal support for leaders to help them better develop their people, the stronger the performance management will be within the organization. The more employees are rewarded—formally and informally—for evaluating performance diligently and (on the

employee side) for actively seeking and incorporating feedback, the more performance management will strongly contribute to organizational performance. The potential impact of I-O psychology on performance management practice is not limited to improved measurement. We have the capacity for enhancing the full lifecycle of performance-relevant behaviors and informing organizations how to build their performance culture. The full gamut of interconnected human capital activities can reinforce that culture: leadership development, employee development, selection and promotion, and rewards and compensation. Therefore, let's stop engaging in the self-deceptive thinking that getting rid of ratings is the key to improving performance management.

Moreover, before we toss out ratings as the villain, we should ask ourselves the broader and much more important question: Are we conducting the entire performance management process properly? Our perspective is that the ratings issue is just representative of deeper problems—a performance management process that is poorly designed and implemented. Our profession has accumulated extensive advice on how to design and implement performance management (e.g., Posthuma & Campion, 2008; Posthuma, Campion, & Campion, 2015), often summarized as "best practices," which if followed correctly may well moot the discussion about ratings.

A Dilemma

In the end, the narrow ratings issue may be a dilemma—a choice between undesirable alternatives. However, that is not new to I-O psychologists. In fact, helping make these hard choices is what we do. Consider the fact that turned-down candidates are not happy and would prefer to do away with our tests and selection systems. Nonpromoted employees would prefer to get rid of our promotion procedures. Employees receiving smaller pay increases will disagree with whatever system was used. So should we abandon performance ratings because managers or employees do not like them and because they are difficult to do well? Will we be better off?

Closing

If managers and employees engaged in effective day-to-day performance management behavior as needed, in real time, there should be less if any need for formal performance management systems, including formal performance ratings. Some managers, in fact, do this and realize high team and individual performance as a result, in spite of the formal performance management system. However, not all or even most managers regularly engage in effective performance management behavior, especially given that the formal system often gets in the way and drives

ineffective performance management behavior and reactions. Although organizations can abandon ratings and even their formal performance management systems entirely, many are not ready to consider such extreme steps. Furthermore, many organizations feel the need to maintain evaluations of record, for legal and other purposes. The merits of these positions have been argued above, but what remains is a practically more important question: How can organizations make the right decisions about performance management reform, including the question of ratings, to best mitigate negative impact on effective performance management behavior and performance?

The concept of performance management is squarely aimed at helping individuals and organizations maximize their productivity through enabling employees to perform to their potential. To achieve this, performance needs to be managed with three critical goals in mind:

- Enable employees to align their efforts to the organization's goals.
- Provide guideposts to monitor behavior and results, and make real-time adjustments to maximize performance.
- Help employees remove barriers to success.

Each aspect of the performance management process should be designed to efficiently and directly impact one or more of these goals. Many organizations seeking to improve their performance management approaches want to start with questions such as "should we have ratings?" or "should we use a forced distribution, and if so what should the cutoff percentage be for the lowest rating?" These are the wrong questions to ask at the start. The better questions to start with are these: What are the critical outcomes we want to achieve, and how can we best ensure employees deliver against key goals and outcomes? Framed from this perspective, there is no right answer to the ratings question. It really is "It depends," based on the organization's goals, strategies, maturity, trust, openness to change, management philosophy, and other contextual factors. Taking a broader and more strategic approach versus a narrow view focused simply on ratings will help keep the focus on what the performance management system needs to attain holistically. This will require answering key questions that matter most in deciding on the right performance management strategy for each situation. We close with those questions organizational members have found most useful for driving effective performance management reform.

- What business problem(s) are we trying to solve?
  - Drive organizational strategy?
  - Reduce unnecessary process and costs?
  - Increase engagement and performance?

- ○ Use formal evaluations to justify differential decisions?
- What do we want to evaluate and reward?
  - ○ Individual or team performance or both?
  - ○ Behaviors, outcomes, both, or other?
- How do we view our employees, and how much do we need to compete for talent?
- What proportion of an employee's total compensation do we put at risk?
- How much do we really rely on ratings for decision making?
  - ○ How much do our ratings actually align with our decisions?
  - ○ Do we need a number to make decisions? Why or why not?
- How mature and ready is our organization to remove formal performance management steps and process?
- How effective are we at the daily behaviors that drive performance?
- Are the answers the same for all units, jobs, and work in the organization?

References

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. *2*, pp. 267–299). New York, NY: Academic Press.

Adler, S. (2015, April). Getting rid of performance ratings: Genius or folly? In S. Adler (Chair), Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

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

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Avoiding a "me" versus "we" dilemma: Using performance management to turn teams into a source of competitive advantage. *Business Horizons, 56*, 503–512. Retrieved from http://mypage.iu.edu/~haguinis/pubs.html

Aguinis, H., Joo, H., & Gottfredson, R. K. (2011). Why we hate performance management—and why we should love it. *Business Horizons, 54*, 503–507. Retrieved from http://mypage.iu.edu/~haguinis/pubs.html

Aguinis, H., & O'Boyle, E., Jr. (2014). Star performers in twenty-first century organizations. *Personnel Psychology, 67*, 313–350.

Aguinis, H., & Pierce, C. A. (2008). Enhancing the relevance of organizational behavior by embracing performance management research. *Journal of Organizational Behavior, 29*, 139–145.

Allen, D. J., & Griffeth, R. W. (2001). Test of a mediated performance–turnover relationship highlighting the moderating roles of visibility and reward contingency. *Journal of Applied Psychology, 86*, 1014–1021.

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology, 47*, 349–356.

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1922. *Journal of Applied Psychology, 77*, 836–874.

Balzer, W. K., & Sulsky, L. M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology, 77*, 975–985.

Banks, C., & Murphy, K. (1985). Toward narrowing the research–practice gap in performance appraisal. *Personnel Psychology, 38*, 335–345.

Barankay, I. (2011). *Rankings and social tournaments: Evidence from a crowd-sourcing experiment* (Working paper). Philadelphia, PA: University of Pennsylvania.

Beck, J. W., Beatty, A. S., & Sackett, P. R. (2014). On the distribution of job performance: The role of measurement characteristics in observed departures from normality. *Personnel Psychology, 67*, 531–566.

Beehr, T. A., Taber, T. D., & Walsh, J. T. (1980). Perceived mobility channels: Criteria for intraorganizational job mobility. *Organizational Behavior and Human Performance, 26*, 250–264.

Benson, P. G., Buckley, R. M., & Hall, S. (1988). The impact of rating scale format on rater accuracy: An appraisal of the mixed standard scale. *Journal of Management, 14*(3), 415–423.

Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent.

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review, 6*, 205–212.

Bernardin, H. J., & Pierce, E. C. (1980). Effect of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology, 65*, 60–66.

Bingham, W. V. (1939). Halo, invalid and valid. *Journal of Applied Psychology, 23*, 221–228.

Blanes i Vidal, J., & Nossol, M. (2009). *Tournaments without prizes: Evidence from personnel records*. Paper presented at the CEP Labour Market Workshop, Centre for Economic Performance, London School of Economics and Political Science, London, United Kingdom.

Blume, B. D., Baldwin, T. T., & Rubin, R. S. (2009). Reactions to different types of forced distribution performance appraisal systems. *Journal of Business Psychology, 24*, 77–91.

Bobko, P., & Colella, A. (1994). Employee reactions to performance standards: A review and research propositions. *Personnel Psychology, 47*, 1–29.

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48*, 587–605.

Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior and Human Performance, 20*, 238–252.

Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology, 64*(4), 410–421.

Bracken, D. W., & Rose, D. S. (2011). When does 360-degree feedback create behavior change? And how would we know it when it does? *Journal of Business & Psychology, 26*, 183–192.

Brockner, J. (1988). *Self-esteem at work*. New York, NY: Lexington Books.

Bryson, B., & Read, J. L. (2009). *Total engagement: Using games and virtual worlds to change the way people work and businesses compete*. Boston, MA: Harvard Business Press.

Buckingham, M., & Goodall, A. (2015, April). Reinventing performance management. *Harvard Business Review*. Retrieved from https://hbr.org/2015/04/reinventing-performance-management

Cadsby, C. B., Song, F., & Tapon, F. (2007). Sorting and incentive effects of pay for performance: An experimental investigation. *Academy of Management Journal, 50*, 387–405.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco, CA: Jossey-Bass.

Campion, M. A., Fink, A. A., Ruggeberg, B. J., Carr, L., Phillips, G. M., & Odman, R. B. (2011). Doing competencies well: Best practices in competency modeling. *Personnel Psychology, 64*, 225–262.

Campion, M. C., Campion, E. D., & Campion, M. A. (2015). Improvements in performance management through the use of 360 feedback. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8*, 85–93.

Carver, C., Sutton, S., & Scheier, M. (2000). Action, emotion, and personality: Emerging conceptual integration. *Personality and Social Psychology Bulletin, 26*, 741–751.

Casas-Arce, P., & Martinez-Jerez, F. A. (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science, 55*, 1306–1320.

Catano, V. M., Darr, W., & Campbell, C. A. (2007). Performance appraisal of behavior-based competencies: A reliable and valid procedure. *Personnel Psychology, 60*, 201–230.

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology, 83*(4), 615–633.

Chamorro-Premuzic, T. (2011). *Personality and individual differences*. London, UK: Blackwell.

Chattopadhayay, R., & Ghosh, A. K. (2012). Performance appraisal based on a forced distribution system: Its drawbacks and remedies. *International Journal of Productivity, 61*(8), 881–896.

Church, A. H. (1995). From both sides now performance appraisals: Political tools or effective measures? *Industrial and Organizational Psychologist, 33*, 57–65.

Cleveland, J. N., & Murphy, K. R. (1992). Analyzing performance appraisal as goal-directed behavior. In G. Ferris & K. Rowland (Eds.), *Research in personnel and human resources management* (Vol. 10, pp. 121–185). Greenwich, CT: JAI Press.

Cleveland, J. N., Murphy, K. R., & Lim, A. (2007). Feedback phobia? Why employees do not want to give or receive it. In J. Langan-Fox, C. Cooper, & R. Klimoski (Eds.), *Research companion to the dysfunctional workplace: Management challenges and symptoms* (pp. 168–186). Cheltenham, UK: Edward Elgar.

Cleveland, J. N., Murphy, K. R., & Williams, R. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology, 74*, 130–135.

Connelly, B. L., Tihanyi, L., Crook, T. R., & Gangloff, K. A. (2014). Tournament theory: Thirty years of contests and competitions. *Journal of Management, 40*, 16–47.

Cooper, W. H. (1981a). Conceptual similarity as a source of illusory halo in job performance ratings. *Journal of Applied Psychology, 66*, 302–307.

Cooper, W. H. (1981b). Ubiquitous halo. *Psychology Bulletin, 90*, 218–244.

Corporate Leadership Council. (2012). *Driving breakthrough performance in the new work environment*. (Catalog No. CLC4570512SYN). Washington, DC: CEB.

Corporate Leadership Council. (2014). *The millennial myth: Three strategies for effectively managing millennials in the workforce*. Washington, DC: CEB.

Culbert, S. A., & Rout, L. (2010). *Get rid of the performance review: How companies can stop intimidating, start managing—and focus on what really matters*. New York, NY: Business Plus.

Culbertson, S. S., Henning, J. B., & Payne, S. C. (2013). Performance appraisal satisfaction: The role of feedback and goal orientation. *Journal of Personnel Psychology, 12*, 189–195.

Cunningham, L. (2014). In big move, Accenture will get rid of annual performance reviews and ranking. *The Washington Post*. Retrieved from 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

Curtis, A. B., Harvey, R. D., & Ravden, D. (2005). Sources of political distortions in performance appraisals: Appraisal purpose and rater accountability. *Group & Organization Management, 30*(1), 42–60.

Day, D. V., & Sulsky, L. M. (1995). Effects of frame-of-reference training and information configuration on memory organization and rating accuracy. *Journal of Applied Psychology, 80*, 158–167.

DeCotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. *Academy of Management Review, 3*, 635–646.

DeNisi, A. S. (2006). *A cognitive approach to performance appraisal*. New York, NY: Routledge.

DeNisi, A. S., Cafferty, T., & Meglino, B. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance, 33*, 360–396.

DeNisi, A. S., & Pritchard, R. D. (2006). Performance appraisal, performance management, and improving individual performance: A motivational framework. *Management and Organization Review, 2*(2), 253–277.

DeNisi, A., & Smith, C. E. (2014). Performance appraisal, performance management, and firm-level performance. *Academy of Management Annals, 8*, 127–179.

De Pater, I. E., Van Vianen, A. E., Bechtoldt, M. N., & Klehe, U. C. (2009). Employees' challenging job experiences and supervisors' evaluations of promotability. *Personnel Psychology, 62*(2), 297–325.

DeVaro, J. (2006). Strategic promotion tournaments and worker performance. *Strategic Management Journal, 27*, 721–740.

Dipboye, R. L., & Pontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. *Journal of Applied Psychology, 66*(2), 248–251.

Dorfman, P. W., Stephan, W. G., & Loveland, J. (1986). Performance appraisal behaviors: Supervisor perceptions and subordinate reactions. *Personnel Psychology, 39*, 579–597.

Dweck, C. (2006). *Mindset*. New York, NY: Random House.

Efron, M., & Ort, M. (2010). *One-page talent management*. Cambridge, MA: Harvard Business Press.

Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology, 86*, 215–227.

Fay, C. H., & Latham, G. P. (1982). Effects of training and rating scales on rating errors. *Personnel Psychology, 35*, 105–116.

Ferris, G. R., Munyon, T. P., Basik, K., & Buckley, M. R. (2008). The performance evaluation context: Social, emotional, cognitive, political, and relationship components. *Human Resource Management Review, 18*, 146–163.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140.

Flint, D. H. (1999). The role of organizational justice in multi-source performance appraisal: Theory-based applications and directions for research. *Human Resource Management Review, 9*(1), 1–20.

Giles, W. F., & Mossholder, K. W. (1990). Employee reactions to contextual and session components of performance appraisal. *Journal of Applied Psychology, 75*(4), 371–377.

Gilliland, S. W., & Langdon, J. C. (1998). Creating performance management systems that promote perceptions of fairness. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 209–243). San Francisco, CA: Jossey-Bass.

Goffin, R. D., Jelley, B. R., Powell, D. M., & Johnston, N. G. (2009). Taking advantage of social comparisons in performance appraisal: The relative percentile method. *Human Resource Management, 48*(2), 251–268.

Grey, R. J., & Kipnis, D. (1976). Untangling the performance appraisal dilemma: The influence of perceived organizational context on evaluative processes. *Journal of Applied Psychology, 61*, 329–335.

Halim, A. H. (2011). Validation on behavioral competencies of a performance appraisal tool. *Journal of Psychological Research, 6*(2), 281–290.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisory, self-peer, and peer-subordinate ratings. *Personnel Psychology, 41*, 43–62.

Harrison, D. A., Virick, M., & William, S. (1996). Working without a net: Time, performance, and turnover under maximally contingent rewards. *Journal of Applied Psychology, 81*, 331–345.

Heneman, H. G. (1974). Comparisons of self- and superior ratings of managerial performance. *Journal of Applied Psychology, 59*, 638–642.

Ilgen, D. R., Barnes-Farrell, J. L., & McKellin, D. B. (1993). Performance appraisal process research in the 1980s: What has it contributed to appraisals in use? *Organizational Behavior and Human Decision Processes, 54*, 321–368.

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. In L. Cummings & B. Staw (Eds.), *Research in organizational behavior* (Vol. 5, pp. 141–197). Greenwich, CT: JAI Press.

Ilgen, D. R., Peterson, R. B., Martin, B. A., & Boeschen, D. A. (1981). Supervisor and subordinate reactions to performance appraisal sessions. *Organizational Behavior and Human Performance, 28*, 311–330.

Jelley, B. R., & Goffin, R. D. (2001). Can performance-feedback accuracy be improved? Effects of rater priming and rating-scale format on rating accuracy. *Journal of Applied Psychology, 86*(1), 134–144.

Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgments of overall performance. *Journal of Applied Psychology, 86*(5), 984–996.

Judge, T. A., & Ferris, G. R. (1993). Social context of performance evaluation decisions. *Academy of Management Journal, 51*, 709–732.

Kluger, A. N., & DeNisi, A. S. (1996). The effects of feedback interventions on performance: A historical review, meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284.

Kozlowski, S. W., Chao, G. T., & Morrison, R. F. (1998). Games raters play: Politics, strategies, and impression management in performance appraisal. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 163–208). San Francisco, CA: Jossey-Bass.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*, 72–107.

Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance*. New York, NY: Academic Press.

Latham, G. P., Almost, J., Mann, S., & Moore, C. (2005). New developments in performance management. *Organizational Dynamics, 34*(1), 77–87.

Latham, G. P., Fay, C. H., & Saari, L. M. (1979). The development of behavioral observation scales for appraising the performance of foremen. *Personnel Psychology, 32*, 299–311.

Latham, G., & Wexley, K. (1977). Behavioral observation scales. *Journal of Applied Psychology, 30*, 255–268.

Lawler, E. E., Benson, G. S., & McDermott, M. (2012). Performance management and reward systems. *WorldatWork Journal, 21*(4), 19–28.

Lazear, E. P. (1986). Salaries and piece rates. *Journal of Business, 59*, 405–431.

London, M., & Wohlers, A. J. (1991). Agreement between subordinate and self-ratings in upward feedback. *Personnel Psychology, 44*, 375–390.

Longenecker, C. O., Sims, H. P., & Gioia, D. A. (1987). Behind the mask: The politics of employee appraisal. *Academy of Management Executive, 1*, 183–193.

McKay, P. F., & McDaniel, M. A. (2006). A reexamination of black-white mean differences in work performance: More data, more moderators. *Journal of Applied Psychology, 91*(3), 538.

Menefee, J., & Murphy, R. (2004). Rewarding and retaining the best: Compensation strategies for top performers. *Benefits Quarterly, 20*, 13–20.

Mercer. (2013). *Global performance management survey report: Executive summary*. Retrieved from http://www.mercer.com/content/dam/mercer/attachments/global/Talent/Assess-BrochurePerfMgmt.pdf

Meyer, H. H. (1980). Self-appraisal of job performance. *Personnel Psychology, 33*, 291–295.

Meyer, H. H., Kay, E., & French, J. (1965). Split roles in performance appraisal. *Harvard Business Review, 43*, 123–129.

Milkovich, G. T., & Wigdor, A. K. (1991). *Pay for performance*. Washington, DC: National Academy Press.

Murphy, K. R. (1982). Difficulties in the statistical control of halo. *Journal of Applied Psychology, 67*, 161–164.

Murphy, K. R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 148–160.

Murphy, K. R., & Anhalt, R. L. (1992). Is halo error a property of the rater, rates, or the specific behaviors observed? *Journal of Applied Psychology, 77*, 494–500.

Murphy, K. R., & Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology, 74*, 619–624.

Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal. An organizational perspective*. Needham Heights, MA: Allyn & Bacon.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-oriented perspectives*. Newbury Park, CA: Sage.

Murphy, K. R., Cleveland, J. N., & Mohler, C. (2001). Reliability, validity, and meaningfulness of multisource ratings. In D. Bracken, C. Timmreck, & A. Church (Eds.), *Handbook of multisource feeback* (pp. 130–148). San Francisco, CA: Jossey-Bass.

Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89*, 158–164.

Murphy, K. R., & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology, 53*, 873–900.

Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology, 78*, 218–225.

Murphy, K. R., Martin, C., & Garcia, M. (1982). Do behavioral observation scales measure observation? *Journal of Applied Psychology, 67*, 562–567.

Nathan, B. R., & Tippins, N. (1989). The consequences of halo "error" in performance ratings: A field study of the moderating effect of halo on test validation results. *Journal of Applied Psychology, 74*, 290–296.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2008). No new terrain: Reliability and construct validity of job performance ratings. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 174–179.

Pfeffer, J. (2007). A modest proposal: How we might change the process and product of managerial research. *Academy of Management Journal, 50*, 1334–1345.

Pichler, S. (2012). The social context of performance appraisal and appraisal reactions: A meta-analysis. *Human Resource Management, 51*(5), 709–732.

Posthuma, R. A., & Campion, M. A. (2008). Twenty best practices for just employee performance reviews. *Compensation & Benefits Review, 40*(1), 47–55.

Posthuma, R. A., Campion, M. C., & Campion, M. A. (2015). *A taxonomic foundation for evidence-based management research on employee performance management.* Manuscript submitted for publication.

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology, 69*, 581–588.

Pulakos, E. D., Mueller-Hanson, R. A., Arad, S., & Moye, N. (2015). Performance management can be fixed: An on-the-job experiential learning approach for complex behavior change. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8*(1), 51–76.

Pulakos, E. D., & O'Leary, R. S. (2011). Why is performance management so broken? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 4*(2), 146–164.

Pulakos, E. D., Schmitt, N., & Ostroff, C. (1986). A warning about the use of a standard deviation across dimensions within ratees to measure halo. *Journal of Applied Psychology, 71*, 29–32.

Reilly, R. R., & McGourty, J. (1998). Performance appraisal in team settings. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 244–275). San Francisco, CA: Jossey-Bass.

Rock, D. (2008). SCARF: A brain-based model for collaborating with and influencing others. *NeuroLeadership Journal, 1*, 1–9.

Rock, D., Davis, J., & Jones, B. (2014, August). Kill your performance ratings. *Strategy + Business*. Retrieved from httpv://www.strategy-business.com/article/00275

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York, NY: Academic Press.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology, 75*, 322–327.

Schleicher, D. J., Bull, R. A., & Green, S. G. (2008). Rater reactions to forced distribution rating systems. *Journal of Management, 35*, 899–927.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Scott, W. R. (1995). *Institutions and organizations.* Thousand Oaks, CA: Sage.

Scullen, S. E., Bergey, P. K., & Aiman-Smith, L. (2005). Forced distribution rating systems and the improvement of workforce potential: A baseline simulation. *Personnel Psychology, 58*, 1–32.

Scullen, S. E., Mount, M. K., & Judge, T. A. (2003). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*, 956–970.

Silverman, S. B., & Wexley, K. N. (1984). Reaction of employees to performance appraisal interviews as a function of their participation in rating scale development. *Personnel Psychology, 37*, 703–710.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*, 149–155.

Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology, 58*, 33–66.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures*. Bowling Green, OH: Author.

Stallings, W. M., & Gillmore, G. M. (1972). Estimating the interjudge reliability of the Ali-Frazier fight. *Journal of Applied Psychology, 56*, 435–436.

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology, 73*, 497–506.

Thompson, D. E., & Thompson, T. A. (1985). Task-based performance appraisal for blue-collar jobs: Evaluation of race and sex effects. *Journal of Applied Psychology, 70*(4), 747–753.

Trank, C. Q., Rynes, S. L., & Bretz, R. D., Jr. (2002). Attracting applicants in the war for talent: Differences in work preferences among high achievers. *Journal of Business and Psychology, 16*, 331–345.

Trevor, C. O., Gerhart, B., & Boudreau, J. W. (1997). Voluntary turnover and job performance: Curvilinearity and the moderating influences of salary growth. *Journal of Applied Psychology, 82*, 44–61.

Tziner, A., & Kopelman, R. E. (2002). Is there a preferred performance rating format? A non-psychometric perspective. *Applied Psychology: An International Review, 51*(3), 479–503.

Tziner, A., & Murphy, K. (1999). Additional evidence of attitudinal influences in performance appraisal. *Journal of Business and Psychology, 13*, 407–419.

Tziner, A., Prince, B., & Murphy, K. (1997). PCPAQ—The questionnaire for measuring the perceived political considerations in performance appraisal: Some new evidence regarding its psychological properties. *Journal of Social Behavior and Personality, 12*, 189–200.

Vance, R. J., Winne, P. S., & Wright, S. E. (1983). A longitudinal examination of rater and ratee effects in performance ratings. *Personnel Psychology, 36*, 609–620.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557–574.

Werner, J. M., & Bolino, M. C. (1997). Explaining U.S. courts of appeals decisions involving performance appraisal: Accuracy, fairness, and validation. *Personnel Psychology, 50*, 1–24.

Wherry, R. J., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology, 35*, 521–555.

Williams, C. R., & Livingstone, L. P. (1994). Another look at the relationship between performance and voluntary turnover. *Academy of Management Journal, 37*, 269–298.